Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2014, Article ID 438291, 12 pages
http://dx.doi.org/10.1155/2014/438291

Research Article

Gradient Learning Algorithms for Ontology Computing

Wei Gao[1] and Linli Zhu[2]

[1] *School of Information and Technology, Yunnan Normal University, Kunming 650500, China*
[2] *School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China*

Correspondence should be addressed to Wei Gao; gaowei@ynnu.edu.cn

Received 7 June 2014; Revised 18 September 2014; Accepted 6 October 2014; Published 29 October 2014

Academic Editor: Karim G. Oweiss

The gradient learning model has been raising great attention in view of its promising perspectives for applications in statistics, data dimensionality reducing, and other specific fields. In this paper, we raise a new gradient learning model for ontology similarity measuring and ontology mapping in multidividing setting. The sample error in this setting is given by virtue of the hypothesis space and the trick of ontology dividing operator. Finally, two experiments presented on plant and humanoid robotics field verify the efficiency of the new computation model for ontology similarity measure and ontology mapping applications in multidividing setting.

1. Introduction and Motivations

The term "ontology" is originally from the field of philosophy and it is used to describe the nature connection of things and the inherent hidden connections of their components. In information and computer science, ontology is a model for knowledge storing and representation and has been widely applied in knowledge management, machine learning, information systems, image retrieval, information retrieval search extension, collaboration, and intelligent information integration. In the past decade, as an effective concept semantic model and a powerful analysis tool, ontology has been widely applied in pharmacology science, biology science, medical science, geographic information system, and social sciences (e.g., see Hu et al., [1], Lambrix and Edberg [2], Mork and Bernstein [3], Fonseca et al., [4], and Bouzeghoub and Elbyed [5]).

The structure of ontology can be expressed as a simple graph. Each concept, object, or element in ontology corresponds to a vertex and each (directed or undirected) edge on an ontology graph represents a relationship (or potential link) between two concepts (objects or elements). Let O be an ontology and G a simple graph corresponding to G. The nature of ontology engineer application can be attributed to get the similarity calculating function which is to compute the similarities between ontology vertices. These similarities represent the intrinsic link between vertices in ontology graph. The goal of ontology mapping is to get the ontology similarity measuring function by measuring the similarity between vertices from different ontologies, such mapping is a bridge between different ontologies, and get a potential association between the objects or elements from different ontologies. Specifically, the ontology similarity function $\mathrm{Sim} : V \times V \to \mathbb{R}^+ \cup \{0\}$ is a semipositive score function which maps each pair of vertices to a nonnegative real number.

Example 1. Ontology technologies are widely used in humanoid robotics in recent years. Different bionic robot has a different structure. Each bionic robot or each component of a bionic robot can be represented as an ontology. Each vertex in ontology stands for a part or a construction, edge between vertices represents a direct physical link between these constructs, or these parts have intrinsic link with its function. Thus, the similarity calculation between vertices in the same ontology allows us to find the degree of association and the potential link between different constructs in bionic robots. Similarity calculation between two different ontologies (i.e., ontology mapping building) allows us to understand the potential association for different components or parts in two biomimetic robots.

Example 2. In information retrieval, ontology concepts are often used in query expansion. The user queries the information related concept A. If we manually set the parameters $M > 0$, the ontology algorithm will find that all concepts B meet $\mathrm{Sim}(A, B) > M$. Then the information related concepts B will be returned to the user as the query expansion for concept A.

Very recently, ontology technologies are employed in a variety of applications. Ma et al. [6] presented a graph derivation representation based technology for stable semantic measurement. Li et al. [7] raised an ontology representation method for online shopping customers knowledge in enterprise information. Santodomingo et al. [8] proposed an innovative ontology matching system that finds complex correspondences by processing expert knowledge from external domain ontologies and in terms of using novel matching tricks. Pizzuti et al. [9] described the main features of the food ontology and some examples of application for traceability purposes. Lasierra et al. [10] argued that ontologies can be used in designing an architecture for monitoring patients at home.

Traditional methods for ontology similarity computation are heuristic and based on pairwise similarity calculation. With high computational complexity and low intuitive , this model requires large parameters selection. One example of traditional ontology similarity computation method is

$$\begin{aligned}\mathrm{Sim}(A, B) = {} & \alpha_1 \mathrm{Sim}_{\mathrm{name}}(A, B) + \alpha_2 \mathrm{Sim}_{\mathrm{instance}}(A, B) \\ & + \alpha_3 \mathrm{Sim}_{\mathrm{attribute}}(A, B) + \alpha_4 \mathrm{Sim}_{\mathrm{structure}}(A, B), \end{aligned} \tag{1}$$

where A and B are two vertices corresponding to two concepts; $0 \le \alpha_1, \alpha_2, \alpha_3, \alpha_4 \le 1$ and $\sum_{i=1}^{4} \alpha_i = 1$; $\mathrm{Sim}_{\mathrm{name}}$, $\mathrm{Sim}_{\mathrm{instance}}$, $\mathrm{Sim}_{\mathrm{attribute}}$, and $\mathrm{Sim}_{\mathrm{structure}}$ are functions of name similarity, instance similarity, attribute similarity, and structure similarity, respectively. These similarity functions are determined by experts directly in terms of their experience. Hence, this model has the following deficiencies:

(i) many parameters rely heavily on the experts;

(ii) high computational complexity and thus being inapplicable to ontology with large number of vertices;

(iii) pairwise similarities fall reflect the ontology structure intuitively.

Thus, a more advanced way to deal with the ontology similarity computation is using ontology learning algorithm which gets an ontology function $f : V \to \mathbb{R}$. By virtue of the ontology function, the ontology graph is mapped into a line which consists of real numbers. The similarity between two concepts then can be measured by comparing the difference between their corresponding real numbers.

The essence of this algorithm is dimensionality reduction. In order to associate the ontology function with ontology application, for vertex v, we use a vector to express all its information (including its name, instance, attribute and structure, and semantic information of the concept which is corresponding to the vertex and that is contained in name and attribute components of its vector). In order to facilitate the representation, we slightly confuse the notations and use v to denote both the ontology vertex and its corresponding vector. The vector is mapped to a real number by ontology function $f : V \to \mathbb{R}$, and the ontology function is a dimensionality reduction operator which maps multidimensional vectors into one-dimensional vectors.

There are several effective methods for getting efficient ontology similarity measure or ontology mapping algorithm in terms of ontology function. Wang et al. [11] considered the ontology similarity calculation in terms of ranking learning technology. Huang et al. [12] raised the fast ontology algorithm in order to cut the time complexity for ontology application. Gao and Liang [13] presented an ontology optimizing model such that the ontology function is determined by virtue of NDCG measure, and it is successfully applied in physics education. Since the large part of ontology structure is the tree, Lan et al. [14] explored the learning theory approach for ontology similarity calculating and ontology mapping in specific setting when the structure of ontology graph has no cycle. In the multidividing ontology setting, all vertices in ontology graph or multiontology graph are divided into k parts corresponding to the k classes of rates. The rate values of all classes are determined by experts. In this way, a vertex in a rate a has larger score than any vertex in rate b (if $1 \le a < b \le k$) under the multidividing ontology function $f : V \to \mathbb{R}$. Finally, the similarity between two ontology vertices corresponding to two concepts (or elements) is judged by the difference of two real numbers which they correspond to. Hence, the multidividing ontology setting is suitable to get a score ontology function for an ontology application if the ontology is drawn into a noncycle structure. Gao and Xu [15] studied the uniform stability of multidividing ontology algorithm and obtained the generalization bounds for stable multidividing ontology algorithms.

In the above described ontology learning algorithms, their optimal ontology function calculation model or its solution strategy is done by gradient calculation. Specifically, the ontology gradient learning algorithm obtains the ontology function vector $\vec{f} = (f^1, f^2, \ldots, f^n)^T$ which maps each vertex into a real number (the value f^i corresponds to vertex v_i). In this sense, it is good or bad policy gradient calculation algorithm that will determine the merits of the ontology algorithm. In this paper, we raise an ontology gradient learning algorithm for ontology similarity measuring and ontology mapping in multidividing setting. The organization of the rest paper is as follows: the notations and ontology gradient computing model are directly presented in Section 2; the detailed description of new ontology algorithms is shown in Section 3; in Section 4, we obtain some theoretical results concerning the sample error and convergence rate; in Section 5, two simulation experiments on plant science and humanoid robotics are designed to test the efficiency of our gradient computation based ontology algorithm, and the data results reveal that our algorithm has high precision ratio for plant and humanoid robotics applications.

2. The Gradient Computation Model for Ontology in Multidividing Setting

In order to combine the machine learning technology and ontology frame, the relevant information for each vertex in ontology graph is represented as an n-dimensional vector. Hence the vertex set V is a subset of $\mathbb{R}^n$ (vertex space or input space for ontology). Assume that V is compact. In the supervised learning, let $Y = \mathbb{R}$ be the label set for V. Denote ρ as a probability measure on $Z = V \times Y$. Let ρ_V and $\rho(\cdot \mid v)$ be the marginal distribution on V and conditional distribution at $v \in V$, respectively. The ontology function $f_\rho : V \to \mathbb{R}$ associated with ρ is described as $f_\rho = \int_Y y d\rho(y \mid v)$.

For each vertex $v \in V$, denote $v = (v^1, v^2, \ldots, v^n)^T \in \mathbb{R}^n$. Then, the gradient of the ontology function f_ρ is the vector of ontology functions

$$\nabla f_\rho = \left(\frac{\partial f_\rho}{f_\rho v^1}, \frac{\partial f_\rho}{f_\rho v^2}, \ldots, \frac{\partial f_\rho}{f_\rho v^m} \right)^T. \tag{2}$$

Let $\mathbf{z} = \{(v_i, y_i)\}_{i=1}^m$ be a random sample independently drawn according to ρ in standard ontology setting. The purpose of standard ontology gradient learning is to learn ∇f_ρ from the sample set $\mathbf{z}$. From the perspective of statistical learning theory, the gradient learning algorithm is based on the Taylor expansion $f_\rho(v) \approx f_\rho(v') + \nabla f_\rho(v')(v - v')$ if two vertices have large common information (i.e., $v \approx v'$). We expect that $y_i \approx f_\rho(v)$ and $y_j \approx f_\rho(u)$ if $v' = v_i'$, $v = v_j$. The demand $v_i \approx v_j$ is met by virtue of setting weights

$$\begin{aligned} w(v) &= w^s(v) = \frac{1}{s^{n+2}} e^{-|v|^2/2v^2}, \\ w_{i,j} &= w_{i,j}^s = \frac{1}{s^{n+2}} e^{-|v_i - v_j|^2/2v^2} = w\left(v_i - v_j\right). \end{aligned} \tag{3}$$

Using unknown ontology function vector $\vec{f} = (f^1, f^2, \ldots, f^n)^T$ to replace ∇f_ρ, then the standard least-square ontology learning algorithm is denoted as

$$\vec{f}_{z,\lambda} = \arg\min_{\vec{f} \in \mathscr{H}_K^n} \left\{ \frac{1}{m^2} \sum_{i,j=1}^n w_{i,j}^s \left(y_i - y_j + \vec{f}(v_i)\left(v_j - v_i\right) \right)^2 + \lambda \left\| \vec{f} \right\|_{\mathscr{H}_K^n}^2 \right\}, \tag{4}$$

where λ and s are two positive constants to control the smoothness of ontology function. Here $K : V \times V \to \mathbb{R}$ is a positive semidefinite, continuous, and symmetric kernel (i.e., Mercer kernel) and $\mathscr{H}_K$ is the reproducing kernel Hilbert space (for short, RKHS) associated with the Mercer kernel K. The notation $\mathscr{H}_K^n$ presented in (4) is the n-fold hypothesis space of $\mathscr{H}_K$ composing of vectors of ontology functions $\vec{f} = (f^1, f^2, \ldots, f^n)^T$ with norm $\|\vec{f}\|_{\mathscr{H}_K^n}^2 = \{\sum_{l=1}^n \|f^l\|_K^2\}^{1/2}$.

By the representation theory in statistical learning theory, the ontology algorithm (4) can be implemented in terms of solving a linear system for the coefficients $\{c_i, \mathbf{z}\}_{i=1}^m$ of $\vec{f}_{\mathbf{z},\lambda} = \sum_{i=1}^m c_{i,\mathbf{z}} K_{v_i}$, where $K_v(v') = K(v, v')$ for $v \in V$ is the ontology function in $\mathscr{H}_K$ and $c_{i,\mathbf{z}} \in \mathbb{R}^n$. Let d be the rank of the matrix $[v_i - v_m]_{i=1}^{m-1}$; hence the coefficient matrix for the linear system has size md. Therefore, this size will become huge if the size of sample set m is large itself. The standard approximation ontology algorithm allows us to solve linear systems with coefficient matrices of smaller sizes.

The gradient learning model for ontology algorithm in standard setting is determined as follows:

$$\begin{aligned} \vec{f}_{t+1}^z = \vec{f}_t^z &- \frac{\eta_t}{m^2} \sum_{i,j=1}^m w_{i,j}^{(s)} \left(y_i - y_j + \vec{f}_t^z(v_i) \cdot \left(v_j - v_i \right) \right) K_{v_i} \\ &- \eta_t \lambda_t \vec{f}_t^z, \end{aligned} \tag{5}$$

where the sample set $\mathbf{z} \in Z^m$, $\vec{f}_1^z = 0$, $t \in \mathbb{Z}$, $\{\eta_t\}$ is the sequence of step sizes and $\{\lambda_t\}$ is the sequence of balance parameters.

For multidividing ontology setting, the vertex in ontology sample set can be divided into k rates. Let $\mathbf{z} = \{\mathbf{z}^1, \mathbf{z}^2, \ldots, \mathbf{z}^k\}$ with $\mathbf{z}^i = \{v_1^i, v_2^i, \ldots, v_{m_k}^i\}$ for $1 \le i \le k$. Denote $|\mathbf{z}_a| = m_a$, $m = \sum_{i=1}^k m_i$ and y_i^a is the label of v_i^a for $1 \le a \le k$ and $1 \le i \le m_a$. Hence, (4) becomes

$$\begin{aligned} &\vec{f}_{\mathbf{z},\lambda} \\ &= \arg\min_{\vec{f} \in \mathscr{H}_K^n} \left\{ \frac{\eta_t}{\sum_{a=1}^{k-1} \sum_{b=i+1}^k m_a m_b} \right. \\ &\qquad \times \sum_{a=1}^{k-1} \sum_{b=i+1}^k \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} w_{\substack{a\,b \\ i\,,j}}^{(s)} \\ &\qquad \left. \times \left(y_i^a - y_j^b + \vec{f}\left(v_i^a\right)\left(v_j^b - v_i^a\right) \right)^2 + \lambda \left\| \vec{f} \right\|_{\mathscr{H}_K^n}^2 \right\}. \end{aligned} \tag{6}$$

We obtain the following gradient computation model for ontology application in multidividing setting which corresponds to (5):

$$\begin{aligned} \vec{f}_{t+1}^{\mathbf{z}} = \vec{f}_t^{\mathbf{z}} &- \frac{\eta_t}{\sum_{a=1}^{k-1} \sum_{b=i+1}^k m_a m_b} \\ &\times \sum_{a=1}^{k-1} \sum_{b=i+1}^k \sum_{i=1}^{m_a} \sum_{j=1}^{m_b} w_{\substack{a\,b \\ i\,,j}}^{(s)} \\ &\times \left(y_i^a - y_j^b + \vec{f}_t^{\mathbf{z}}\left(v_i^a\right) \cdot \left(v_j^b - v_i^a \right) \right) K_{v_i^a} - \eta_t \lambda_t \vec{f}_t^{\mathbf{z}}. \end{aligned} \tag{7}$$

Here in (6) and (7), $w_{\substack{a\,b \\ i\,,j}}^{(s)} = (1/s^{n+2}) e^{-((v_i^a)^2 - (v_j^b)^2)/2s^2}$.

We emphasize that our algorithm in multidividing setting is different from that of Wu et al. [16]. First, the label y for ontology vertex v is used to present its class information in [16], that is, $y \in \{1, \ldots, k\}$, while in our setting, $y \in \mathbb{R}$.

Second, the computation model in [16] relies heavily on the convexity loss function l, while our algorithm depends on the weight function w.

3. Description of Ontology Algorithms via Gradient Learning

The above raised gradient learning ontology algorithm can be used in ontology concepts similarity measurement and ontology mapping. The basic idea is the following: via the ontology gradient computation model, the ontology graph is mapped into a real line consisting of real numbers. The similarity between two concepts then can be measured by comparing the difference between their corresponding real numbers.

Algorithm 3 (gradient calculating based ontology similarity measure algorithm). For $v \in V(G)$ and f is an optimal ontology function determined by gradient calculating, we use one of the following methods to obtain the similar vertices and return the outcome to the users.

Method 1. Choose a parameter U and return set $\{v' \in V(G), |f(v') - f(v)| \le U\}$.

Method 2. Choose an integer U and return the closest N concepts on the value list in $V(G)$.

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

Algorithm 4 (gradient calculating based ontology mapping algorithm). Let $G_1, G_2, \ldots, G_d$ be ontology graphs corresponding to ontologies $O_1, O_2, \ldots, O_d$. For $v \in V(G_i)$ $(1 \le i \le d)$ and f being an optimal ontology function determined by gradient calculating, we use one of the following methods to obtain the similar vertices and return the outcome to the users.

Method 1. Choose a parameter U and return set $\{v' \in V(G - G_i), |f(v') - f(v)| \le U\}$.

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

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

4. Theoretical Analysis

In this section, we give certain theoretical analysis for our proposed multidividing ontology algorithm. Let $\kappa = \sup_{v \in V} \sqrt{K(v,v)}$ and $\mathrm{Diam}(V) = \sup_{v,v' \in V} |v - v'|$. We divide this section into two parts: first, some useful lemmas are prepared; then, main results in our paper concerning approximation conclusions are presented. Our error analysis depends on integral operators and gradient learning, and more references on these tricks can be referred to in Mukherjee and Wu [18], Mukherjee et al. [19], Yao et al. [20], and Rosasco et al. [21].

Set

$$\begin{aligned}
&\vec{f}_\lambda^{\,*} \\
&= \arg\min_{\vec{f} \in \mathcal{H}_K^n} \left\{ \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \int_{Z_a} \int_{Z_b} w\left(v^a - v^b\right) \right. \\
&\qquad \times \left(y^a - y^b + \vec{f}\,(v)\right. \\
&\qquad \left. \left. \cdot \left(v^b - v^a\right)^2\right)^2 d\rho \right. \\
&\qquad \times \left(v^a, y^a\right) d\rho\left(v^b, y^b\right) \\
&\qquad \left. + \lambda \left\| \vec{f} \right\|_{\mathcal{H}_K^n}^2 \right\}.
\end{aligned} \tag{8}$$

In what follows, $m_\Pi = m_1 m_2 \cdots m_k$,

$$\begin{aligned}
m_\Pi^1 &= \frac{m_\Pi}{m_1} = m_2 m_3 \cdots m_k, \\
m_\Pi^2 &= \frac{m_\Pi}{m_2} = m_1 m_3 \cdots m_k, \\
&\vdots, \\
m_\Pi^k &= \frac{m_\Pi}{m_k} = m_1 m_2 \cdots m_{k-1}.
\end{aligned} \tag{9}$$

Our tricks of proofs in this paper follow from [22, 23].

4.1. Preliminary Results. Let sequence $\{\vec{f}_t\}_{t \in \mathbb{N}}$ be the noise-free limit of the sequence (7) which is determined by $\vec{f}_1 = 0$ and

$$\begin{aligned}
&\vec{f}_{t+1} \\
&= \vec{f}_t - \eta_t \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \int_{Z^a} \int_{Z^b} w\left(v^a - v^v\right) \\
&\qquad \times \left(y^a - y^b + \vec{f}_t\,(v)\right. \\
&\qquad \left. \cdot \left(v^b - v^a\right)\right) \\
&\qquad \times \left(v^b - v^a\right) K_v d\rho\left(v^a, y^a\right) d\rho\left(v^b, y^b\right) \\
&\qquad - \eta_t \lambda_t \vec{f}_t.
\end{aligned} \tag{10}$$

Our error analysis for proving main result (Theorems 12 and 13 in the next subsection) consists of two parts: sample error and approximation error.

The main task in this subsection is to estimate the sample error $\|\vec{f}_t^{\,\mathbf{z}} - \vec{f}_t\|$ in terms of McDiarmid-Bernstein-type probability inequality and the multidividing sampling operator. For each $1 \le a \le k$, the multidividing sampling operator $S_v^a : \mathcal{H}_K^n \to \mathbb{R}^{m_a n}$ associated with a discrete subset $\mathbf{v}^a = \{v_i^a\}_{i=1}^{m_a}$ of V is defined by

$$S_{\mathbf{v}}^a\left(\vec{f}\right) = \left(\vec{f}\left(v_i^a\right)\right)_{i=1}^{m_a} = \left(\vec{f}\left(v_1^a\right), \vec{f}\left(v_2^a\right), \ldots, \vec{f}\left(v_{m_a^a}\right)\right)^T. \tag{11}$$

The adjoint of the multidividing ontology sampling operator, $(S_{\mathbf{v}}^a)^T : \mathbb{R}^{m_a n} \to \mathscr{H}_K^n$, is given by

$$(S_{\mathbf{v}}^a)^T (c) = \sum_{i=1}^{m_a} c_i^a K_{v_i^a}, \tag{12}$$

where

$$c = (c_i)_{i=1}^{m_a} = \left(c_1, c_2, \ldots, c_{m_a}\right)^T \in \mathbb{R}^{m_a n}. \tag{13}$$

Let us express (7) by virtue of the multidividing ontology sampling operator. Note that

$$\begin{aligned} &\vec{f}_t^{\mathbf{z}}(v_i^a) \cdot \left(v_j^b - v_i^a\right)\left(v_j^b - v_i^a\right) \\ &\quad = \left(v_j^b - v_i^a\right)\left(v_j^b - v_i^a\right)^T \vec{f}_t^{\mathbf{z}}(v_i^a) \\ &\quad = \left(v_j^b - v_i^a\right)\left(v_j^b - v_i^a\right)^T \left(S_v\left(\vec{f}_t^{\mathbf{z}}\right)\right)_i^a. \end{aligned} \tag{14}$$

For each pair of (a, b) with $1 \le a < b \le k$, we single out one summation $\sum_{j=1}^{m_b}$ from (7) as

$$\begin{aligned} B_i^{a,b} &= \sum_{j=1}^{m_b} w_{\substack{a\,b \\ i,j}} \left(v_j^b - v_i^a\right)\left(v_j^b - v_i^a\right)^T \in \mathbb{R}^{n\times n}, \\ Y_i^{a,b} &= \sum_{j=1}^{m_b} w_{\substack{a\,b \\ i,j}} \left(y_j^b - y_i^a\right)\left(v_j^b - v_i^a\right)^T \in \mathbb{R}^{n}. \end{aligned} \tag{15}$$

We infer that

$$\begin{aligned} \vec{f}_{t+1}^{\mathbf{z}} = {} & \left(1 - \eta_t\lambda_t\right)\vec{f}_t^{\mathbf{z}} - \frac{\eta_t}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \\ & \times \left\{-\sum_{i=1}^{m_a}\sum_{j=1}^{m_b} Y_i^{a,b} K_{v_i^a} + \sum_{i=1}^{m_a}\sum_{j=1}^{m_b} K_{v_i^a} B_i^{a,b} \vec{f}_t^{\mathbf{z}}(v_i^a)\right\}. \end{aligned} \tag{16}$$

Denote

$$\begin{aligned} \left(D_{\mathbf{v}}^a\right)^{a,b} &= \operatorname{diag}\left\{B_1^{a,b}, B_2^{a,b}, \ldots, B_{m_a}^{a,b}\right\} \in \mathbb{R}^{m_a n \times m_a n} \\ \left(\vec{Y}^a\right)^{a,b} &= \left(Y_1^{a,b}, Y_2^{a,b}, \ldots, Y_{m_a}^{a,b}\right)^T \in \mathbb{R}^{m_a n}. \end{aligned} \tag{17}$$

Hence, we have

$$\begin{aligned} \vec{f}_{t+1}^{\mathbf{z}} = {} & \left(1 - \eta_t\lambda_t\right)\vec{f}_t^{\mathbf{z}} + \frac{\eta_t}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \\ & \times \left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \left(S_{\mathbf{v}}^a\right)^T \left(\left(\vec{Y}^a\right)^{a,b}\right)^T\right\} \\ & - \frac{\eta_t}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \\ & \times \left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \left(S_{\mathbf{v}}^a\right)^T \left(D_{\mathbf{v}}^a\right)^{a,b} S_{\mathbf{v}}^a\left(\vec{f}_t^{\mathbf{z}}\right)\right\}. \end{aligned} \tag{18}$$

Thus, it confirms the following representation for the sequence $\{\vec{f}_t^{\mathbf{z}}\}$. For simplicity, let $\prod_{q=t+1}^{t}(I - L_{\mathbf{v},q}) = I$ in the following contents.

Lemma 5. *Set*

$$\begin{aligned} L_{\mathbf{v},t} = {} & \frac{\eta_t}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \left(S_{\mathbf{v}}^a\right)^T \left(D_{\mathbf{v}}^a\right)^{a,b} S_{\mathbf{v}}^a\right\} \\ & + \eta_t\lambda_t I. \end{aligned} \tag{19}$$

If $\{\vec{f}_t^{\mathbf{z}}\}$ is defined by (7), we deduce

$$\begin{aligned} \vec{f}_t^{\mathbf{z}} = {} & \Pi_{i=1}^{t-1}\left(I - L_{\mathbf{v},i}\right)\vec{f}_1^{\mathbf{z}} \\ & + \sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left(I - L_{\mathbf{v},q}\right)\frac{\eta_t}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \\ & \times \left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \left(S_{\mathbf{v}}^a\right)^T \left(\left(\vec{Y}^a\right)^{a,b}\right)^T\right\}. \end{aligned} \tag{20}$$

We should discuss the convergence of the multidividing ontology operator

$$\frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \left(S_{\mathbf{v}}^a\right)^T \left(D_{\mathbf{v}}^a\right)^{a,b} S_{\mathbf{v}}^a\left(\vec{f}_t^{\mathbf{z}}\right)\right\} \tag{21}$$

to the integral operator $L_{K,s} : \mathscr{H}_K^n \to \mathscr{H}_K^n$ determined by

$$\begin{aligned} L_{K,s}\vec{f} = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \Bigg\{\int_{V^a}\int_{V^b} w\left(v^a - v^b\right)\left(v^b - v^a\right)\left(v^b - v^a\right)^T \\ \cdot \vec{f}\left(v^a\right) K_{v^a} d\rho_{V^a}\left(v^a\right) d\rho_{V^b}\left(v^b\right)\Bigg\}, \end{aligned} \tag{22}$$

where $\vec{f} \in \mathscr{H}_K^n$.

Lemma 6. *Let $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\}$ be multidividing sample set independently drawn according to a probability distribution ρ on Z. Denote $(H, \|\cdot\|)$ as a Hilbert space and suppose that $F : Z^{m_1\times m_2\times\cdots\times m_k} \to \mathscr{H}$ is measurable. If there is nonnegative $\widetilde{M}$ such that $\|F(\mathbf{z}) - \mathbb{E}z(F(\mathbf{v}))\| \le \widetilde{M}$ for each $v \in \mathbf{z}$ and almost every $\mathbf{z} \in Z^{m_1\times m_2\times\cdots\times m_k}$, then for every $\varepsilon > 0$,*

$$\begin{aligned} & \mathbb{P}_{\mathbf{z}\in Z^{m_1\times m_2\times\cdots\times m_k}}\left\{\left\|F(\mathbf{z}) - \mathbb{E}_{\mathbf{z}}(F(\mathbf{z}))\right\| \ge \varepsilon\right\} \\ & \qquad \le 2\exp\left\{-\frac{\varepsilon^2}{2\left(\widetilde{M}\varepsilon + \sigma^2\right)}\right\}, \end{aligned} \tag{23}$$

where

$$\sigma^2 = \sum_{a=1}^{k}\sum_{i=1}^{m_a} \sup_{\mathbf{z}\setminus\{v_i^a\}\in Z^{m_1\times m_2\times\cdots\times(m_a-1)\times\cdots\times m_k}} \mathbb{E}_{v_i}\left\{\left\|F(\mathbf{z}) - \mathbb{E}_{v_i}(F(\mathbf{z}))\right\|^2\right\}. \tag{24}$$

For any $0 < \delta < 1$, with confidence $1 - \delta$, one gets

$$\begin{aligned} \left\|F(\mathbf{z}) - \mathbb{E}_{\mathbf{z}}(F(\mathbf{z}))\right\| &\le 4\log\frac{4}{\delta}\left\{\widetilde{M} + \sqrt{\sigma^2}\right\} \\ &\le 4\left(1 + \sqrt{\frac{m_\Pi}{\sum_{i=1}^{k} m_\Pi^i}}\right)\log\frac{4}{\delta}\widetilde{M}. \end{aligned} \tag{25}$$

By regarding $(1/\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b)\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}(S_{\mathbf{v}}^a)^T (D_{\mathbf{v}}^a)^{a,b} S_{\mathbf{v}}^a(\vec{f}_t^{\mathbf{z}})\}$ and $L_{K,s}$ as elements in $(L(\mathscr{H}_K^n)$ and $\|\cdot\|_{L(\mathscr{H}_K^n)}$, the space of bounded linear multidividing ontology operators on $\mathscr{H}_K^n$, Lemma 6 cannot be directly employed because $L(\mathscr{H}_K^n)$ is not a Hilbert space, but a Banach space only. Therefore, we consider a subspace of $L(\mathscr{H}_K^n)$, $HS(\mathscr{H}_K^n)$ which is the space of Hilbert-Schmidt operators on $\mathscr{H}_K^n$ with inner product $\langle A, B\rangle_{HS(\mathscr{H}_K^n)} = Tr(B^T A)$. As $HS(\mathscr{H}_K^n)$ is a subspace of $L(\mathscr{H}_K^n)$, their norm relations are presented as

$$\begin{aligned} \|A\|_{L(\mathscr{H}_K^n)} &\le \|A\|_{HS(\mathscr{H}_K^n)}, \\ \|AB\|_{HS(\mathscr{H}_K^n)} &\le \|A\|_{HS(\mathscr{H}_K^n)} \|B\|_{HS(\mathscr{H}_K^n)}. \end{aligned} \tag{26}$$

In addition, $HS(\mathscr{H}_K^n)$ is a Hilbert space and contains multidividing ontology operators $L_{K,s}$ and $(1/\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b)\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}(S_{\mathbf{v}}^a)^T (D_{\mathbf{v}}^a)^{a,b} S_{\mathbf{v}}^a(\vec{f}_t^{\mathbf{z}})\}$. By applying Lemma 6 to this Hilbert space, we obtain the following lemma.

Lemma 7. *Let $\mathbf{v} = \{\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^k\}$ be multidividing sample set independently drawn from (V, ρ_V). With confidence $1-\delta$, one obtains*

$$\begin{aligned} &\left\| \frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}} \left(S_v^{a,b}\right)^T D_v S_v^{a,b} - L_{K,s} \right\|_{HS(\mathscr{H}_K^n)} \\ &\quad \le \frac{34\sqrt{n}\kappa^2 (\text{ Diam }(V))^2}{\sqrt{m_\Pi / \sum_{i=1}^{k} m_\Pi^i}\, s^{n+2}} \log\frac{4}{\delta}. \end{aligned} \tag{27}$$

Proof. Let $H = HS(H_K^n)$. Consider the multidividing ontology function $F : V^{m_1\times m_2\times\cdots\times m_k} \to \mathscr{H}$ with values in $H = HS(\mathscr{H}_K^n)$ defined by

$$F(\mathbf{v}) = \frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \left\{ \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} (S_{\mathbf{v}}^a)^T (D_{\mathbf{v}}^a)^{a,b} S_{\mathbf{v}}^a \right\}. \tag{28}$$

For $\vec{f} \in \mathscr{H}_K^n$, we confirm that

$$\begin{aligned} &F(\mathbf{v})\left(\vec{f}\right) \\ &= \frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}} \sum_{i=1}^{m_a}\sum_{j=1}^{m_b} w\left(v_i^a - v_j^b\right)\left(v_j^b - v_i^a\right) \\ &\qquad \times \left(v_j^b - v_i^a\right)^T \vec{f}\left(v_i^a\right) K_{v_i^a}. \end{aligned} \tag{29}$$

Recall that reproducing property of the RKHS $\mathscr{H}_K$ says that

$$f(v) = \langle f, K_v\rangle_K, \quad \forall v \in V,\ f \in \mathscr{H}_K. \tag{30}$$

It implies that the rank of operator $A_v : \mathscr{H}_K \to \mathscr{H}_K$ determined by $A_v(f) = f(v)K_v = \langle f, K_v\rangle_K K_v$ is 1, and also in $HS(\mathscr{H}_K)$. Furthermore, $\|A_v\|HS(\mathscr{H}_K) = K(v,v)$. Let $\vec{A}_v$ be the operator on $\mathscr{H}_K^n$ which maps $\vec{f}$ to $\vec{f}(v)K_v$. Then the above fact reveals that $\|\vec{A}_v\|_{HS(\mathscr{H}_K^n)} \le K(v,v)\sqrt{n}$. Hence for any $\mathbf{v} \in V^{m_1\times m_2\times\cdots\times m_k}$, we infer that

$$\begin{aligned} F(\mathbf{v}) = \frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}} &\sum_{i=1}^{m_a}\sum_{j=1}^{m_b} w\left(v_i^a - v_j^b\right)\left(v_j^b - v_i^a\right) \\ &\times \left(v_j^b - v_i^a\right)^T \vec{A}_{v_i^a} \in HS\left(\mathscr{H}_K^n\right). \end{aligned} \tag{31}$$

Using the fact that $w(v) \le 1/s^{n+2}$ and $\|\vec{A}_v\|_{HS(\mathscr{H}_K^n)} \le \sqrt{n}K(v,v) \le \sqrt{n}\kappa^2$, we deduce that

$$\begin{aligned} &\left\| F(\mathbf{v}) - \mathbb{E}_{v_i}(F(\mathbf{v})) \right\|_{HS(\mathscr{H}_K^n)} \\ &\quad \le \frac{4\left(m_\Pi / \sum_{i=1}^{k} m_\Pi^i - 1\right)\kappa^2 (\text{Diam}\,(V))^2\sqrt{n}}{\left(m_\Pi / \sum_{i=1}^{k} m_\Pi^i\right)^2 s^{n+2}}. \end{aligned} \tag{32}$$

Since

$$\begin{aligned} &\mathbb{E}_{\mathbf{v}}\left(\frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \left\{ \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} (S_{\mathbf{v}}^a)^T (D_{\mathbf{v}}^a)^{a,b} S_{\mathbf{v}}^a \right\} \right) \\ &\quad = \mathbb{E}_{\mathbf{v}}(F(\mathbf{v})) = \frac{m_\Pi / \sum_{i=1}^{k} m_\Pi^i}{m_\Pi / \sum_{i=1}^{k} m_\Pi^i - 1} L_{K,s}, \end{aligned} \tag{33}$$

the stated result is held by combining Lemma 6 with

$$\widetilde{M} = (\text{Diam}\,(V))^2 \kappa^2 \sqrt{n} \frac{8\left(m_\Pi / \sum_{i=1}^{k} m_\Pi^i - 1\right)}{\left(m_\Pi / \sum_{i=1}^{k} m_\Pi^i\right)^2 s^{n+2}} \tag{34}$$

and using the bound $\|L_{K,s}\|_{HS(\mathscr{H}_K^n)} \le \kappa^2\sqrt{n}(\text{Diam}(V))^2/s^{n+2}$. □

In order to find the difference between $\vec{f}_t^{\mathbf{z}}$ and $\vec{f}_t$, the convergence of

$$\frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \left\{ \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} (S_{\mathbf{v}}^a)^T \left(\left(\vec{Y}^a\right)^{a,b} \right)^T \right\} \tag{35}$$

to the ontology function defined by (55) is studied.

Lemma 8. *Let $\mathbf{z}$ be a multidividing ontology sample independently drawn from (Z, ρ). With confidence $1-\delta$, one has*

$$\begin{aligned} &\left\| \frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b} \left\{ \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} (S_{\mathbf{v}}^a)^T \left(\left(\vec{Y}^a\right)^{a,b} \right)^T \right\} - \vec{f}_{\rho,s} \right\|_{\mathscr{H}_K^n} \\ &\quad \le \frac{68\ \text{Diam}\ (V)\, M\kappa}{\sqrt{m_\Pi / \sum_{i=1}^{k} m_\Pi^i}\, s^{n+2}} \log\frac{4}{\delta}. \end{aligned} \tag{36}$$

Proof. By applying Lemma 6 to the Hilbert space $H = \mathcal{H}_K^n$ and the ontology function $F : Z^{m_1\times m_2\times\cdots\times m_k} \to \mathcal{H}_K^n$ given by

$$\begin{aligned} &F(\mathbf{z}) \\ &= \frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b}\left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left(S_{\mathbf{v}}^a\right)^T\left(\left(\vec{Y}^a\right)^{a,b}\right)^T\right\} \\ &= \frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}}\sum_{i=1}^{m_a}\sum_{j=1}^{m_b} w\left(v_i^a - v_j^b\right)\left(v_j^b - v_i^a\right)\left(v_j^b - v_i^a\right)K_{v_i^a}, \end{aligned} \tag{37}$$

we yield $\mathbb{E}_{\mathbf{z}}(F(\mathbf{z})) = ((m_\Pi/\sum_{i=1}^k m_\Pi^i - 1)/m_\Pi/\sum_{i=1}^k m_\Pi^i)\vec{f}_{\rho,s}$. Hence, for almost every $\mathbf{z} \in Z^{m_1\times m_2\times\cdots\times m_k}$, we get

$$\begin{aligned} &\left\|F(\mathbf{z}) - \mathbb{E}_{z_i}(F(\mathbf{z}))\right\|_{\mathcal{H}_K^n} \\ &\quad\le \frac{16M\kappa\mathrm{Diam}(V)\left(m_\Pi/\sum_{i=1}^k m_\Pi^i - 1\right)}{\left(m_\Pi/\sum_{i=1}^k m_\Pi^i\right)^2 s^{n+2}}. \end{aligned} \tag{38}$$

Lemma 6 implies that for any $0 < \delta < 1$, with confidence $1-\delta$, we obtain

$$\begin{aligned} &\left\|\frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b}\left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left(S_{\mathbf{v}}^a\right)^T\left(\left(\vec{Y}^a\right)^{a,b}\right)^T\right\}\right. \\ &\quad\left. - \frac{m_\Pi/\sum_{i=1}^k m_\Pi^i - 1}{m_\Pi/\sum_{i=1}^k m_\Pi^i}\vec{f}_{\rho,s}\right\|_{\mathcal{H}_K^n} \\ &\quad\le \frac{32\left(1 + 1/\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}\right)M\kappa\mathrm{Diam}(V)}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}\,s^{n+2}}\log\frac{4}{\delta}. \end{aligned} \tag{39}$$

Finally, conclusion follows from the fact that $\|\vec{f}_{\rho,s}\|_{\mathcal{H}_K^n} \le 4\mathrm{Diam}(V)M\kappa/s^{n+2}$. □

Obviously, for $\{\vec{f}_t^{\mathbf{z}}\}$, the sequence $\{\vec{f}_t\}$ has a similar expression as (20).

Lemma 9. *Let $L_{K,\lambda_i,\eta_i} = \eta_i L_{K,s} + \eta_i\lambda_i I$ be an ontology operator on $\mathcal{H}_K^n$ and suppose that $\prod_{q=i+1}^{t-1}(I - L_{K,\lambda a_k,\eta_k}) = I$. For the ontology operator $L_{K,s}$ determined by (22) and $\{\vec{f}_t\}$ by (10), one obtains*

$$\vec{f}_t = \prod_{i=1}^{t-1}\left(I - L_{K,\lambda_i,\eta_i}\right)\vec{f}_1 + \sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left(I - L_{K,\lambda_k,\eta_k}\right)\eta_i\vec{f}_{\rho,s}. \tag{40}$$

The sample error $\|\vec{f}_t^{\mathbf{z}} - \vec{f}_t\|_{\mathcal{H}_K^n}$ is stated in the following conclusion.

Theorem 10. *Let $\{\vec{f}_t^{\mathbf{z}}\}$ be obtained by (5) and $\{\vec{f}_t\}$ by (10). Suppose that $\eta_i \le 1$ and $\lambda_{i+1} \le \lambda_i \le 1$ for all $i \in \mathbb{N}$. Then for any $0 < \delta < 1$, with confidence $1-\delta$, one infers that*

$$\begin{aligned} &\left\|\vec{f}_t^{\mathbf{z}} - \vec{f}_t\right\|_{\mathcal{H}_K^n} \\ &\quad\le \frac{34\ \mathrm{Diam}\ (V)\kappa}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}\lambda_{t-1}^2 s^{n+2}} \\ &\qquad\times\left\{\kappa\sqrt{n}\ \mathrm{Diam}\ (V) + 4\lambda_{t-1}M\right\}\log\frac{8}{\delta}. \end{aligned} \tag{41}$$

Proof. Let

$$\vec{f}_{\rho,t}^{\mathbf{z}} = \sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left(I - L_{\mathbf{v},k}\right)\eta_i\vec{f}_{\rho,s} + \prod_{i=1}^{t-1}\left(I - L_{\mathbf{v},i}\right)\vec{f}_1^{\mathbf{z}}. \tag{42}$$

Let $Z_1 \subseteq Z^{m_1\times m_2\times\cdots\times m_k}$ with measure at least $1-\delta$ such that (36) establishes for any $\mathbf{z} \in Z_1$. Thus, from the positivity of the multidividing ontology operator $(S_{\mathbf{v}}^a)^T(D_{\mathbf{v}}^a)^{a,b}S_{\mathbf{v}}^a$ (for each pair of (a,b)) on $\mathcal{H}_K^n$ and the assumption $\prod_{q=t+1}^{t-1}(1-\eta_q\lambda_q) = 1$, we have that for any $\mathbf{z} \in Z_1$,

$$\begin{aligned} &\left\|\vec{f}_t^{\mathbf{z}} - \vec{f}_{\rho,t}^{\mathbf{z}}\right\|_{\mathcal{H}_K^n} \\ &= \left\|\sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left(I - L_{\mathbf{v},q}\right)\eta_i\right. \\ &\qquad\times\left(\frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} m_a m_b}\right. \\ &\qquad\left.\left.\times\left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left(S_{\mathbf{v}}^a\right)^T\left(\left(\vec{Y}^a\right)^{a,b}\right)^T\right\} - \vec{f}_{\rho,s}\right)\right\|_{L(\mathcal{H}_K^n)} \\ &\le \sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left\|I - L_{\mathbf{v},k}\right\|_{L(\mathcal{H}_K^n)}\frac{68\ \mathrm{Diam}\ (V)M\kappa}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}\,s^{n+2}}\log\frac{4}{\delta} \\ &\le \frac{68\ \mathrm{Diam}\ (V)M\kappa}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}\,s^{n+2}}\log\frac{4}{\delta}\sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left(1-\eta_q\lambda_q\right)\eta_i. \end{aligned} \tag{43}$$

In terms of $\eta_i\lambda_i = 1 - (1-\eta_i\lambda_i)$ and $1 \le \lambda_i\lambda_{t-1}^{-1}$, we get

$$\begin{aligned} &\sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left(1-\eta_q\lambda_q\right)\eta_i \\ &\quad\le \frac{1}{\lambda_{t-1}}\left\{\sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left(1-\eta_q\lambda_q\right) - \sum_{i=1}^{t-1}\prod_{q=i}^{t-1}\left(1-\eta_q\lambda_q\right)\right\} \\ &\quad= \frac{1}{\lambda_{t-1}}\left\{1 - \prod_{q=1}^{t-1}\left(1-\eta_q\lambda_q\right)\right\}. \end{aligned} \tag{44}$$

By virtue of the assumptions on η_i, λ_i, we infer that

$$\sum_{i=1}^{t-1}\prod_{q=i+1}^{t-1}\left(1-\eta_q\lambda_q\right)\eta_i \le \frac{1}{\lambda_{t-1}}, \tag{45}$$

which implies that

$$\left\|\vec{f}_t^{\mathbf{z}}-\vec{f}_{\rho,t}^{\mathbf{z}}\right\|_{\mathscr{H}_K^n} \le \log\frac{4}{\delta}\frac{68\ \mathrm{Diam}\ (V)\,M\kappa}{s^{n+2}\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}\lambda_{t-1}} \tag{46}$$

for any $\mathbf{z}\in Z_1$.

Now, we consider the estimate of $\|\vec{f}_t^{z}-\vec{f}_{\rho,t}^{\mathbf{z}}\|_{\mathscr{H}_K^n}$. Let $Z_2\subseteq Z^{m_1\times m_2\times\cdots\times m_k}$ with measure at least $1-\delta$ such that (27) is established for any $\mathbf{z}\in Z_2$. In view of (26), for each $\mathbf{z}\in Z_2$ we yield

$$\begin{aligned}&\left\|\frac{1}{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}m_am_b}\left\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left(S_{\mathbf{v}}^a\right)^T\left(D_{\mathbf{v}}^a\right)^{a,b}S_{\mathbf{v}}^a\right\}-L_{K,s}\right\|_{L(\mathscr{H}_K^n)}\\ &\quad\le\log\frac{2}{\delta}\frac{34\sqrt{n}\kappa^2(\ \mathrm{Diam}\ (V))^2}{s^{n+2}\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}}.\end{aligned} \tag{47}$$

Using the fact that $L_{K,\lambda_j,n_j}-L_{\mathbf{v},j}=\eta_j(L_{K,s}-(1/\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}m_am_b)\{\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}(S_{\mathbf{v}}^a)^T(D_{\mathbf{v}}^a)^{a,b}S_{\mathbf{v}}^a\})$, we obtain that for any $\mathbf{z}\in Z_2$,

$$\begin{aligned}&\left\|\vec{f}_t-\vec{f}_{\rho,t}^{z}\right\|_{\mathscr{H}_K^n}\\ &=\left\|\sum_{i=1}^{t-1}\left(\prod_{q=i+1}^{t-1}\left(I-L_{\mathbf{v},q}\right)-\prod_{l=i+1}^{t-1}\left(I-L_{K,\lambda_l,n_j}\right)\right)\eta_i\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}\\ &=\left\|\sum_{i=1}^{t-1}\sum_{j=i+1}^{t-1}\prod_{q=j+1}^{t-1}\left(I-L_{\mathbf{v},q}\right)\left(L_{K,\lambda_j,n_j}-L_{\mathbf{v},q}\right)\right.\\ &\qquad\left.\times\prod_{l=i+1}^{t-1}\left(I-L_{K,\lambda_l,n_j}\right)\eta_i\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}\\ &\le\sum_{i=1}^{t-1}\sum_{j=i+1}^{t-1}\prod_{q=j+1}^{t-1}\left(1-\eta_q\lambda_q\right)\eta_j\\ &\quad\times\frac{17\kappa^2(\ \mathrm{Diam}\ (V))^2\sqrt{n}}{\sqrt{m_\Pi/\sum_{i=1}m_i'}s^{n+2}}\log\frac{2}{\delta}\prod_{l=i+1}^{j-1}\left(1-\eta_l\lambda_l\right)\eta_i\left\|\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}.\end{aligned} \tag{48}$$

By changing the order of summation, we determine that

$$\begin{aligned}&\left\|\vec{f}_t-\vec{f}_{\rho,t}^{z}\right\|_{\mathscr{H}_K^n}\\ &\quad\le\frac{34\kappa^2(\ \mathrm{Diam}\ (V))^2\sqrt{n}}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}s^{n+2}}\log\frac{4}{\delta}\\ &\qquad\cdot\sum_{j=2}^{t-1}\prod_{q=j+1}^{t-1}\left(1-\eta_q\lambda_q\right)\eta_j\\ &\qquad\times\sum_{i=1}^{j-1}\prod_{l=i+1}^{j-1}\left(1-\eta_l\lambda_l\right)\eta_i\left\|\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}.\end{aligned} \tag{49}$$

According to (45), we can verify that $\|\vec{f}_t-\vec{f}_{\rho,t}^{\mathbf{z}}\|_{\mathscr{H}_K^n}$ is bounded by

$$\begin{aligned}&\log\left(\frac{4}{\delta}\right)\frac{34\kappa^2(\ \mathrm{Diam}\ (V))^2\sqrt{n}}{\sqrt{m_\Pi/\sum_{i=1}^k m_{\Pi_i}}s^{n+2}}\\ &\qquad\times\sum_{j=2}^{t-1}\prod_{q=j+1}^{t-1}\left(1-\eta_q\lambda_q\right)\eta_j\frac{1}{\lambda_{j-1}}\left\|\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}\\ &\quad\le\log\left(\frac{4}{\delta}\right)\frac{34\kappa^2(\ \mathrm{Diam}\ (V))^2\sqrt{n}}{\sqrt{m_\Pi/\sum_{i=1}^k m_{\Pi_i}}s^{n+2}}\frac{1}{\lambda_{j-1}^2}\left\|\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}.\end{aligned} \tag{50}$$

In view of the above fact and (46), we obtain that for any $\mathbf{z}\in Z_1\cap Z_2$,

$$\begin{aligned}&\left\|\vec{f}_t-\vec{f}_{\rho,t}^{\mathbf{z}}\right\|_{\mathscr{H}_K^n}\\ &\quad\le\log\frac{2}{\delta}\left\{\frac{68\kappa^2(\ \mathrm{Diam}\ (V))^2\sqrt{n}}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}s^{n+2}}\right.\\ &\qquad\left.+\frac{34\kappa^2(\ \mathrm{Diam}\ (V))^2\sqrt{n}}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}s^{n+2}}\left\|\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}\right\}.\end{aligned} \tag{51}$$

However, the measure of the subset $Z_1\cap Z_2$ of $Z^{m_1\times m_2\times\cdots\times m_k}$ is at least $1-2\delta$. The desired conclusion follows after substituting δ for $\delta/2$. □

The following result is Theorem 4 in Dong and Zhou [23]; it also holds in multidividing setting and we skip the detailed proof.

Theorem 11. *Let $\{\lambda_t,\eta_t\}_{t\in\mathbb{N}}$ be determined by* (53). *Then, we deduce that*

$$\begin{aligned}&\left\|\vec{f}_t-\vec{f}_{\lambda_t}^*\right\|_{\mathscr{H}_K^n}\\ &\le t^{2\gamma+\alpha-1}\left\{4\gamma C_{\lambda_1\eta_1,\gamma+\alpha,1-\gamma}+\exp\left\{\frac{\lambda_1\eta_1-\log\left(e\lambda_1\eta_1\right)}{1-\gamma-\alpha}\right\}\right\}\\ &\quad\times\frac{\left\|\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}}{\lambda_1}.\end{aligned} \tag{52}$$

4.2. Main Results. The first main result in our paper implies that $\{\vec{f}_t^{\mathbf{z}}\}$ is a good approximation of a noise-free limit for the ontology function (6) as a solution of (8) which we refer as multidividing ontology function $\vec{f}_\lambda^*$.

Theorem 12. *Let $0 < \gamma, \alpha < 1$, and λ_1 and $\eta_1 > 0$ satisfy $2\gamma + \alpha < 1$ and $\lambda_1\eta_1 < 1$. For any $t \in \mathbb{N}$, take*

$$\lambda_t = \lambda_1 t^{-\alpha}. \tag{53}$$

Define $\{\vec{f}_t^{\mathbf{z}}\}$ by (7) and $\{\vec{f}_\lambda^*\}$ by (8). If $|y| \le M$ is almost established, then for any $0 < \delta < 1$, with confidence $1 - \delta$, one has

$$\left\|\vec{f}_t^{\mathbf{z}} - \vec{f}_{\lambda_t}^*\right\|_{\mathscr{H}_K^n} \le \widetilde{C}\left\{\log\frac{8}{\delta}\frac{t^{2\gamma}}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}s^{n+2}} + t^{2\gamma+\alpha-1}\right\} \times\left(1 + \left\|\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n}\right), \tag{54}$$

where constant $\widetilde{C}$ independent of $m_1, m_2, \ldots, m_k, t, s$ or δ and $\vec{f}_{\rho,s}$ is the multidividing ontology function determined by

$$\vec{f}_{\rho,s} = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\int_{V^a}\int_{V^b} w_{a,b}^{(s)}\left(v^a - v^b\right)\left(f_\rho\left(v^b\right) - f_\rho\left(v^a\right)\right) \times\left(v^b - v^a\right)K_v d\rho_V\left(v^a\right)d\rho_V\left(v^b\right). \tag{55}$$

The proof of Theorem 12 follows from Theorems 10 and 11 and an exact expression for the constant $\widetilde{C}$ relying on $\alpha, \eta_1, \lambda_1, \kappa, n, \gamma, M$ and Diam (V) *can be easily determined.*

The second main result in our paper follows from Theorem 10 and the technologies raised in [23].

Theorem 13. *Assume that for certain $0 < \tau \le 2/3$, $c_\rho > 0$ and for any $s > 0$, the marginal distribution ρ_V satisfies*

$$\rho_V\left(\left\{v \in V : \inf_{u\in\mathbb{R}^n\setminus V}|u - v| \le s\right\}\right) \le c_\rho^2 s^{4s}, \tag{56}$$

and the density $p(v)$ of $d\rho_V(v)$ exists and for any, any $u, v \in V$ satisfies

$$\sup_{v\in V} p(v) \le c_\rho, \qquad \left|p(v) - p(u)\right| \le c_\rho|u - v|^\tau. \tag{57}$$

Suppose that the kernel $K \in C^3$ and $\nabla f_\rho \in \mathscr{H}_K^n$. Let $0 < \beta < 1/(4+(2n+4)\gamma/\tau)$ and $0 < \gamma < 2/5$. Take $\lambda_t = t^{-\gamma}$, $\eta_t = t^{(5/2)\gamma-1}$, and $s = s(m_1, m_2, \ldots, m_k) = (\kappa c_\rho)^{2/\tau}\left(m_\Pi/\sum_{i=1}^k m_\Pi^i\right)^{-\beta\gamma/\tau}$ and suppose that $\left(m_\Pi/\sum_{i=1}^k m_\Pi^i\right)^\beta \le t \le 2\left(m_\Pi/\sum_{i=1}^k m_\Pi^i\right)^\beta$; then for any $0 < \delta < 1$, with confidence $1 - \delta$, one infers that

$$\left\|\vec{f}_t^{\mathbf{z}} - \nabla f_\rho\right\|_{(L_{\rho_V^2})^n} \le \widetilde{C}_{\rho,K}\left(\frac{1}{m_\Pi/\sum_{i=1}^k m_\Pi^i}\right)^\theta \log\left(\frac{4}{\delta}\right), \tag{58}$$

where

$$\theta = \min\left\{\frac{1}{2} - 2\beta - \frac{(n+2)\beta\gamma}{\tau}, \frac{\beta\gamma}{2}\right\} \tag{59}$$

and constant $\widetilde{C}_{\rho,K}$ is independent of $m_1, m_2, \ldots, m_k, t$ or δ.

Proof. Obviously, under the assumptions $K \in C^3$, (56) and (57), we get

$$\left\|\vec{f}_{\rho,s}\right\|_{\mathscr{H}_K^n} \le C_{\rho,K}\left(c_\rho n(2\pi)^{n/2}\kappa^2\left\|\nabla f_\rho\right\|_{\mathscr{H}_K^n} + s\right). \tag{60}$$

Furthermore, by virtue of Proposition 15 in Mukherjee and Zhou [22], we have

$$\left\|\vec{f}_t^* - \nabla f_\rho\right\|_{(L_{\rho_V^2})^n} \le C_{\rho,K}\left\{\left\|\nabla f_\rho\right\|_{\mathscr{H}_K^n}\sqrt{\lambda} + \frac{s}{\lambda}\right\}, \tag{61}$$

where constant $C_{\rho,K}$ relies on ρ and K. Theorem 10 and these estimates reveal that with confidence $1 - \delta$, we yield

$$\left\|\vec{f}_t^{\mathbf{z}} - \nabla f_\rho\right\|_{(L_{\rho_V^2})^n} \le \widetilde{C}'\left(1 + \left\|\nabla f_\rho\right\|_{\mathscr{H}_K^n}\right) \times\log\frac{4}{\delta}\left\{\frac{t^{2\gamma}}{\sqrt{m_\Pi/\sum_{i=1}^k m_\Pi^i}s^{n+2}} + t^{2\gamma+\alpha-1} + st^\gamma + t^{-\gamma/2}\right\}. \tag{62}$$

The learning rate (58) is determined according to the selection of the parameters. □

5. Experiments

To show the effectiveness of our new ontology algorithms, two experiments concerning ontology measure and ontology mapping are designed below.

5.1. Ontology Similarity Measure Experiment on Plant Data. In the first experiment, we use plant "PO" ontology O_1 which was constructed in the website http://www.plantontology.org/. The structure of O_1 is presented in Figure 1. *P@N* (precision ratio; see Craswell and Hawking [24]) is used to measure the quality of the experiment data. Here, we take $k = 2$, $t = 3$, $\eta_t = 1$, and $\lambda = 0.1$.

We first give the closest N concepts for every vertex on the ontology graph by experts in plant field, and then we obtain the first N concepts for every vertex on ontology graph by Algorithm 3 and compute the precision ratio. Specifically, for vertex v and given integer $N > 0$. Let $\mathrm{Sim}_v^{N,\mathrm{expert}}$ be the set

Table 1: The experiment results of ontology similarity measure.

	$P@3$ average precision ratio	$P@5$ average precision ratio	$P@10$ average precision ratio
Algorithm 3 in our paper	0.5042	0.6216	0.7853
Algorithm in [11]	0.4549	0.5117	0.5859
Algorithm in [12]	0.4282	0.4849	0.5632
Algorithm in [13]	0.4831	0.5635	0.6871

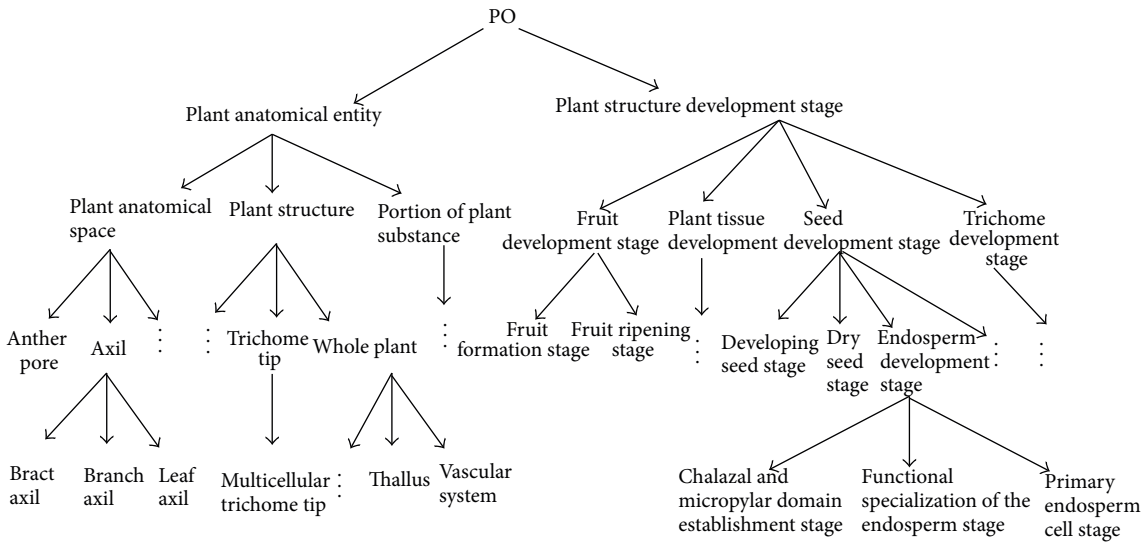

Figure 1: The structure of "PO" ontology.

of vertices determined by experts and it contains N vertices having the most similarity of v. Let

$$\begin{aligned} v_v^1 &= \arg\min_{v' \in V(G)-v} \left\{ \left| f(v) - f\left(v'\right) \right| \right\}, \\ v_v^2 &= \arg\min_{v' \in V(G)-\{v, v_v^1\}} \left\{ \left| f(v) - f\left(v'\right) \right| \right\}, \\ &\vdots \\ v_v^N &= \arg\min_{v' \in V(G)-\{v, v_v^1, \ldots, v_v^{N-1}\}} \left\{ \left| f(v) - f\left(v'\right) \right| \right\}, \\ \mathrm{Sim}_v^{N,\mathrm{algorithm}} &= \left\{ v_v^1, v_v^2, \ldots, v_v^N \right\}. \end{aligned} \tag{63}$$

Then the precision ratio for vertex v is denoted by

$$\mathrm{Pre}_v^N = \frac{\left| \mathrm{Sim}_v^{N,\mathrm{algorithm}} \cap \mathrm{Sim}_v^{N,\mathrm{expert}} \right|}{N}. \tag{64}$$

The $P@N$ average precision ratio for ontology graph G is then stated as

$$\mathrm{Pre}_G^N = \frac{\sum_{v \in V(G)} \mathrm{Pre}_v^N}{|V(G)|}. \tag{65}$$

At the same time, we apply ontology methods in [11–13] to the "PO" ontology. Calculating the average precision ratio by these three algorithms and comparing the results to Algorithm 3 rose in our paper, part of the data is referred to in Table 1.

When $N = 3$, 5, or 10, the precision ratio by virtue of our gradient computation based algorithm is higher than the precision ratio determined by algorithms proposed in [11–13]. In particular, when N increases, such precision ratios are increasing apparently. Therefore, the gradient learning based ontology Algorithm 3 described in our paper is superior to the method proposed by [11–13].

5.2. Ontology Mapping Experiment on Humanoid Robotics Data. For the second experiment, we use "humanoid robotics" ontologies O_2 and O_3. The structure of O_2 and O_3 is shown in Figures 2 and 3, respectively. The ontology O_2 presents the leg joint structure of bionic walking device for six-legged robot, while the ontology O_3 presents the exoskeleton frame of a robot with wearable and power-assisted lower extremities. In this experiment, we take $k = 2$, $t = 4$, $\eta_t = 1$, and $\lambda = 0.05$.

The goal of this experiment is to give ontology mapping between O_2 and O_3. We also use $P@N$ precision ratio to measure the quality of experiment. Again, we apply ontology algorithms in [12, 13, 17] on "humanoid robotics" ontology and compare the precision ratio which is gotten from three methods. Some results referred to in Table 2.

Taking $N = 1$, 3, or 5, the precision ratio in terms of our gradient computation based ontology mapping algorithm is higher than the precision ratio determined by algorithms

Table 2: The experiment results of ontology mapping.

	$P@1$ average precision ratio	$P@3$ average precision ratio	$P@5$ average precision ratio
Algorithm 4 in our paper	0.4444	0.5185	0.6111
Algorithm in [17]	0.2778	0.4815	0.5444
Algorithm in [12]	0.2222	0.4074	0.4889
Algorithm in [13]	0.2778	0.4630	0.5333

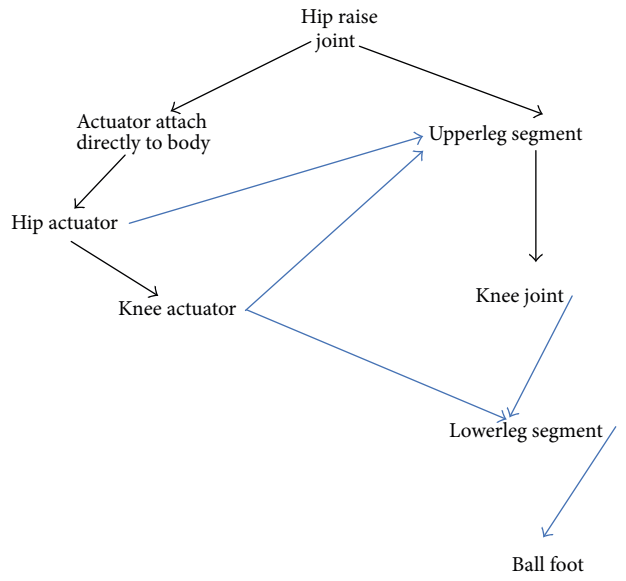

Figure 2: "Humanoid robotics" ontology O_2.

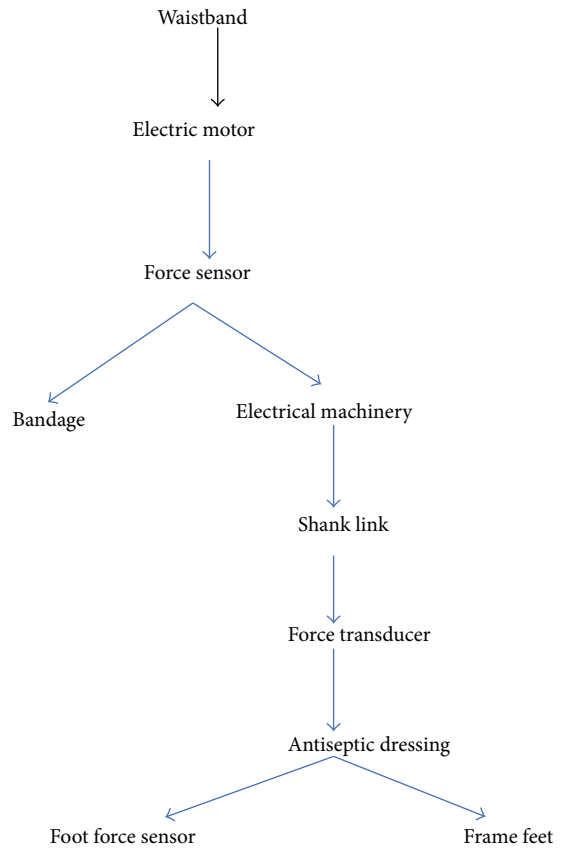

Figure 3: "Humanoid robotics" ontology O_3.

proposed in [12, 13, 17]. Particularly, as N increases, the precision ratios in view of our algorithm are increasing apparently. Therefore, the gradient learning based ontology Algorithm 4 described in our paper is superior to the method proposed by [12, 13, 17].

6. Conclusions

As a data structural representation and storage model, ontology has been widely used in various fields and proved to have a high efficiency. The core of ontology algorithm is to get the similarity measure between vertices on ontology graph. One learning trick is mapping each vertex to a real number, and the similarity is judged by the difference between the real number which the vertices correspond to. In this paper, we raise a gradient learning model for ontology application in multidividing setting. The sample error and approximation properties are given in our paper. These results support the gradient computation based ontology algorithm from the theoretical point of view. The new technology contributes to the state of the art for applications and the result achieved in our paper illustrates the promising application prospects for multidividing ontology algorithm.

Conflict of Interests

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

Acknowledgments

This work was supported in part by the Key Laboratory of Educational Informatization for Nationalities, Ministry of Education, the National Natural Science Foundation of China (60903131), the College Natural Science Foundation of Jiangsu Province in China (10KJD520002), and the Ph.D. initial funding of the first author. The authors are grateful to the anonymous referee for careful checking of the details and for helpful comments that improved this paper.

References

[1] B. Hu, S. Dasmahapatra, P. Lewis, and N. Shadbolt, "Ontology-based medical image annotation with description logics," in *Proceedings of the 15th IEEE International Conference on Tools with artificial Intelligence*, pp. 77–82, Sacramento, Calif, USA, November 2003.

[2] P. Lambrix and A. Edberg, "Evaluation of ontology tools in bioinformatics," in *Pacific Symposium on Biocomputing*, pp. 529–600, IEEE Computer Society Press, New York, NY, USA, 2003.

[3] P. Mork and P. A. Bernstein, "Adapting a generic match algorithm to align ontologies of human anatomy," in *Proceedings of the 20th International Conference on Data Engineering (ICDE '04)*, pp. 787–790, IEEE Computer Society Press, New York, NY, USA, April 2004.

[4] F. Fonseca, M. 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, no. 1-2, pp. 121–151, 2002.

[5] A. Bouzeghoub and A. Elbyed, "Ontology mapping for web-based educational systems interoperability," *Interoperability in Business Information Systems*, vol. 1, no. 1, pp. 73–84, 2006.

[6] Y. Ma, L. Liu, K. Lu, B. Jin, and X. Liu, "A graph derivation based approach for measuring and comparing structural semantics of ontologies," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 5, pp. 1039–1052, 2014.

[7] Z. Li, S. S. Guo, Y. S. Yuan, and L. B. Sun, "Ontology representation of online shopping customers knowledge in enterprise information," *Applied Mechanics and Materials*, vol. 483, pp. 603–606, 2014.

[8] R. Santodomingo, S. Rohjans, M. Uslar, J. A. Rodríguez-Mondéjar, and M. A. Sanz-Bobi, "Ontology matching system for future energy smart grids," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 242–257, 2014.

[9] T. Pizzuti, G. Mirabelli, M. A. Sanz-Bobi, and F. Goméz-Gonzaléz, "Food track & trace ontology for helping the food traceability control," *Journal of Food Engineering*, vol. 120, no. 1, pp. 17–30, 2014.

[10] N. Lasierra, A. Alesanco, and J. Garcia, "Designing an architecture for monitoring patients at home: ontologies and web services for clinical and technical management integration," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 896–906, 2014.

[11] Y. Y. Wang, W. Gao, Y. G. Zhang, and Y. Gao, "Ontology similarity computation use ranking learning method," in *Proceedings of the 3rd International Conference on Computational Intelligence and Industrial Application*, pp. 20–22, Wuhan, China, 2010.

[12] X. Huang, T. W. Xu, W. Gao, and Z. Y. Jia, "Ontology similarity measure and ontology mapping via fast ranking method," *International Journal of Applied Physics and Mathematics*, vol. 1, no. 1, pp. 54–59, 2011.

[13] 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.

[14] M. H. Lan, J. Xu, and W. Gao, "Ontology similarity computation using k-partite ranking method," *Journal of Computer Applications*, vol. 32, no. 4, pp. 1094–1096, 2012.

[15] W. Gao and T. Xu, "Stability analysis of learning algorithms for ontology similarity computation," *Abstract and Applied Analysis*, vol. 2013, Article ID 174802, 9 pages, 2013.

[16] J. Z. Wu, X. Yu, and W. Gao, "Learning rate of gradient descent multi-dividing ontology algorithm," submitted to *International Journal of Manufacturing Technology and Management*.

[17] W. Gao and M. H. Lan, "Ontology mapping algorithm based on ranking learning method," *Microelectronics and Computer*, vol. 28, no. 9, pp. 59–61, 2011.

[18] S. Mukherjee and Q. Wu, "Estimation of gradients and coordinate covariation in classification," *Journal of Machine Learning Research*, vol. 7, pp. 2481–2514, 2006.

[19] S. Mukherjee, Q. Wu, and D.-X. Zhou, "Learning gradients on manifolds," *Bernoulli*, vol. 16, no. 1, pp. 181–207, 2010.

[20] Y. Yao, L. Rosasco, and A. Caponnetto, "On early stopping in gradient descent learning," *Constructive Approximation*, vol. 26, no. 2, pp. 289–315, 2007.

[21] L. Rosasco, M. Belkin, and E. de Vito, "On learning with integral operators," *Journal of Machine Learning Research*, vol. 11, pp. 905–934, 2010.

[22] S. Mukherjee and D.-X. Zhou, "Learning coordinate covariances via gradients," *Journal of Machine Learning Research*, vol. 7, pp. 519–549, 2006.

[23] X. Dong and D.-X. Zhou, "Learning gradients by a gradient descent algorithm," *Journal of Mathematical Analysis and Applications*, vol. 341, no. 2, pp. 1018–1027, 2008.

[24] N. Craswell and D. Hawking, "Overview of the TREC 2003 web track," in *Proceedings of the 12th Text Retrieval Conference*, NIST Special Publication, pp. 78–92, Gaithersburg, Md, USA, 2003.