Analysis of gradient descent ontology iterative algorithm for geological setting

Hualong Wu, Bo Zhao and Wei Gao

School of Information Science and Technology, Yunnan Normal University, Kunming, China

1. Introduction

The term "Ontology" comes from philosophy and used to describe the nature of things. In computer science, ontology is defined as a shared conceptual model, which has been applied in image retrieval, knowledge management, information retrieval search extension, information systems, collaboration, and intelligent information integration. Ontology, as an effective concept semantic model, has been widely employed in many other fields such as education science (Bouzeghoub & Elbyed, 2006), biology medicine (Hu, Dasmahapatra, Lewis, & Shadbolt, 2003), and geography science (Fonseca, Egenhofer, Davis, & Camara, 2001).

Each vertex on an ontology graph represents a concept; each edge on an ontology graph represents a connection between two concepts. Let \( G \) be a graph corresponding to ontology \( O \); the aim of ontology similarity measure is to approach a similarity function which maps each vertex to a real number. The similarity between two vertices is measured by the difference of their corresponding scores. We propose the gradient descent ontology algorithm for this setting, and the learning rates with general convex losses are determined.

be given by optimizing NDCG measure, and applied such idea in physics education. Gao and Gao (2012) obtained the ontology function using the regression approach. Huang, Xu, Gao, and Gong (2011) presented ontology algorithm based on half transductive learning. Lan, Xu, and Gao (2012) explored the learning theory approach for ontology similarity computation in a setting when the ontology graph is a tree. Lin, Weige, Xiaohong, and Wei (2016) proposed new criterion for multi-dividing ontology algorithm from AUC standpoint, which was designed to avoid the choice of loss function. Gao, Gao, and Liang (2013) presented new algorithms for ontology similarity measurement and ontology mapping in terms of harmonic analysis and diffusion regularization on hypergraph. Recently, Gao and Shi (2013) proposed new algorithms for ontology similarity measurement such that the new computational models consider operational cost in the real implement.

Several papers have contributed to the theoretical analysis of different ontology settings. Gao and Xu (2013) investigated the uniform stability of multi-dividing ontology algorithm and gave the generalization bounds for stable multi-dividing ontology algorithms. Gao, Gao, and Zhang (2012) researched strong and weak stabilities of multi-dividing ontology algorithm. Gao and Xu (2012) learned some characteristics for such ontology algorithm. Gao, Xu, Gan, and Zhou (2014) studied the multi-dividing ontology algorithm from a theoretical view. It is highlighted that empirical multi-dividing ontology model can be expressed as conditional linear statistical, and an approximation result is achieved based on projection method. Gao, Gao, and Liang (2016) presented the characteristics of the best ontology score function among piece constant ontology score functions.
Gao, Gao, Zhang, and Liang (2014) investigated the upper bound and the lower bound minimax learning rate is obtained based on low noise assumptions. Gao, Yan, and Liang (2014) and Yan, Gao, and Li (2013) presented an approach of piecewise constant function approximation for AUC criterion multi-dividing ontology algorithms.

In this paper, we consider the ontology problem using ranking learning technology. The gradient descent ontology algorithm is designed, and the error rates for gradient learning with general convex losses are presented. The organization of this paper is as follows: we introduce the basic setting in Section 2; in Section 3, we present several preliminary lemmas; and in Section 4, we prove the main result in this paper.

2. Setting

We use a vector to express all the information for each vertex in ontology graph. Then, we say V takes its value in a high dimension feature space. Let \( V \) and \( Y \subseteq \mathbb{R} \) be, respectively, an input and an output space; \( V \) is called instance space which instances in \( V \) drawn randomly and independently according to some (unknown) distribution \( \rho \). Given a training set \( S = \{(v_1,y_1), \ldots, (v_m,y_m)\} \) of size \( m \) in \( V \times Y \), it is usually assumed that the ontology algorithm \( A \) is symmetric with respect to \( S \). The goal of ontology algorithm is to obtain an ontology (score) function \( f : V \rightarrow \mathbb{R} \), which assigns each vertex a score.

For \( v, v' \in V \), the order of \( v \) and \( v' \) under ontology function \( f \) should be consistent with their labels. The ontology loss function \( l : \mathbb{R}^V \times (V \times Y) \times (V \times Y) \rightarrow \mathbb{R}^+ \cup \{0\} \) is used to punish inconsistency situation which sgn

\[
(l(v), l(v')) = \text{sgn}(y - y'), \text{where } \text{sgn}(x) = \begin{cases} 
1 & x > 0 \\
0 & x = 0 \\
-1 & x < 0 
\end{cases}
\]

\( l \) also satisfies that \( l(f(v), y), (v, y') \) is a non-negative real number and symmetric with respect to \( (v, y) \) and \( (v', y') \).

The quality of ontology function is measured by the expected l-error:

\[
R_l(f) = \sum \left( l(f(v), y), (v', y') \right) \rho(v, y) \rho(v', y').
\]

However, it can’t be estimated directly since distribution \( \rho \) is unknown. We use empirical l-error to measure ontology algorithm instead:

\[
\hat{R}_l(f, S) = \frac{1}{m} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} l(f(v_i, y_i), (v_j, y_j)).
\]

However, the ontology function given by minimum \( \hat{R}_l(f, S) \) usually has bad smoothness. It often adds a regularization term to measure the problem, and use regularized empirical l-error with regularization parameter \( \lambda > 0 \) to measure \( f : \hat{R}_l(f, S) + \lambda N(f) \) where \( F \) is a class of real-valued functions on \( V \), and a regularization function \( N : F \rightarrow \mathbb{R}^+ \cup \{0\} \).

Ontology algorithm that, given a training sample \( S \), outputs a ontology function \( f_S \in F \) satisfies:

\[
f_S = \arg \min_{f \in F} \hat{R}_l(f, S).
\]

Let \( F \) be the measurable ontology function space and \( \mathcal{F} = \{ f \in F : f = \arg \min_{f \in F} R_l(f) \} \) be the collection of optimal ontology functions. Note that the optimal ontology function is not unique. A Mercer kernel is a symmetric and positive semi-definite continuous function \( K : V \times V \rightarrow \mathbb{R} \). The RKHS \( H_K \) associated with the kernel \( K \) is defined to be the closure of the linear span of the set of functions \( \{ K(v, \cdot) : v \in V \} \) with the inner product \( \langle \cdot, \cdot \rangle_K \) given by \( \langle K(v, K(y)) = K(v, y) \). Let

\[
\kappa = \sup_{v \in V} \sqrt{K(v, v)}.
\]

For any \( v \in V \) and \( f \in H_K \), we have

\[
|f(v)| \leq \sqrt{K(v, v)} \| f \|_{H_K} \text{ and } \| f \|_{\infty} \leq \kappa \| f \|_{H_K}.
\]

The regularized ontology algorithm associated with RKHS can be implemented by

\[
f_{\lambda, \delta} = \arg \min_{f \in H_K} \{ \hat{R}_l(f, S) + \lambda \| f \|_{H_K}^2 \}.
\]

The expected version of Equation (1) is

\[
f_{\lambda} = \arg \min_{f \in H_K} \{ R_l(f) + \lambda \| f \|_{H_K}^2 \}.
\]

The aim of this paper is to present the statistic property of gradient descent ontology algorithm which is to learn an optimal ontology function in \( \mathcal{F} \). In terms of the convexity of \( l \), we deduce that its left derivative \( l' \) is well defined and non-decreasing on \( \mathbb{R} \).

Let \( t \in \mathbb{N} \) and \( \eta_t \) be the sequence of step sizes. The stochastic gradient descent ontology algorithm is denoted for the sample \( S \) by \( f_t = 0 \) and

\[
f_{t+1} = (1 - \eta_t \lambda) f_t - \frac{\eta_t}{m} \sum_{i=1}^{m} l'(f(v_i, y_i), (v_i, y_i))(K_{v_i} - K_{y_i}).
\]

Our main tasks are to analyze the errors \( \| f_t^* - f_t \|_{H_K} \) and inf\( \min_{f \in \mathcal{F}} \| f_t^* - f \|_{H_K} \). Suppose that \( l \in \mathcal{C}(\mathbb{R}) \) with

\[
|l(u)| \leq C_0(1 + |u|^q)l'(u) \leq C_0(1 + |u|^{q-1}), \text{for any } u \in \mathbb{R}
\]

where \( C_0 > 0 \) and \( q \geq 1 \). Denote the constant

\[
\Delta_* = 1 + 4\kappa^2(\sup_{|u| \leq 1} |l'(u) - l'(0)| + C_0 l'(0)) + (8\kappa^2 l'(0))^{q-1}.
\]

Following are the main conclusions in our paper.

**Theorem 1:** Suppose ontology loss function \( l \) satisfies (4), and the step size is chosen as

\[
\eta_t = \eta_t \max(0, 1) \quad \text{for some } 0 < \theta < 1, 0 < \eta_t \leq \frac{1}{\Delta_*}
\]
For $0 < y < \min\{q + 1, 1\}$ and $\epsilon > 0$, take $\lambda = \epsilon y$ with $\min\{q - 1, 1\} \leq \epsilon \leq 2 \min\{q - 1, 1\}$. Then, for any $0 < \epsilon < 1$, with confidence at least $1 - \delta$, we get

$$\left\| f^\lambda - f_0 \right\|_{H_k}^2 \leq C \left( \frac{m}{2} \right)^{1 - \epsilon},$$

(7)

where $C$ is a constant independent of $m$, and

$$\alpha = \min\{s \beta - sy \min\{q + 1, 2q - 1\}, 1 - sy(1 + q)\}.$$  

(8)

Using Theorem 1, we derive the estimates of $\inf_{f \in \theta} \left\| f^\lambda - f \right\|_{H_k}$ under some approximation conditions.

**Corollary 1:** Suppose that there exist $f \in \Theta$ with $\left\| f^\lambda - f \right\|_{H_k} \leq C_\lambda \lambda^p$ for some $0 < \beta < 1$. Under the condition in Theorem 1, for any $0 < \epsilon < 1$, with confidence at least $1 - \delta$, we obtain

$$\inf_{f \in \theta} \left\| f^\lambda - f \right\|_{H_k}^2 \leq \hat{C} \left( \frac{m}{2} \right)^{1 - \alpha}$$

(9)

where $\hat{C}$ is a constant independent of $m$, and

$$\hat{\alpha} = \min\{s \beta - sy \min\{q + 1, 2q - 1\}, 1 - sy(1 + q), s \beta\}.$$  

(10)

### 3. Preliminary lemmas

The main contributions of this section are to deduce some preliminary lemmas for proofing Theorem 1. Our first lemma determines a special character of $R(f) + \frac{\lambda^2}{2} \left\| f \right\|_{H_k}^2$. Since the proof is similar as that in (Ying & Zhou, 2006), we leave the detail proof to the readers.

**Lemma 1:** Let $\lambda > 0$ and $f \in H_k$. We yield

$$\frac{\lambda}{2} \left\| f^\lambda - f \right\|_{H_k} \leq \left\{ R(f) + \frac{\lambda}{2} \left\| f \right\|_{H_k}^2 \right\}$$

(11)

By virtue of Lemma 1, we derive the one-step analysis.

**Lemma 2:** Let $\varphi(S, t) = R(f_0) - R(f_0) + R(f_0^\lambda) - R(f_0^\lambda, S)$. For $t \geq 0$, we infer

$$\left\| f^\lambda_{t+1} - f_0 \right\|_{H_k}^2 \leq (1 - \eta_t \lambda) \left\| f^\lambda_t - f_0 \right\|_{H_k}^2 + \eta_t^2 \left\| f^\lambda_t \right\|_{H_k}^2 + 2\eta_t \varphi(S, t).$$

(12)

**Proof:** In view of the convexity of $l$ and the Schwartz inequality, we verify

$$\left\| f^\lambda_{t+1} - f_0 \right\|_{H_k}^2 = \left\| f^\lambda_t - f_0 \right\|_{H_k}^2 + \eta_t^2 \left\| f^\lambda_t \right\|_{H_k}^2 + 2\eta_t \left( f^\lambda_t - f_0, f^\lambda_t - f_0 \right)_K$$

(13)

and

$$\left\langle f_0 - f^\lambda_t, f^\lambda_t \right\rangle_K = \frac{\eta_t}{m} \sum_{j=1}^m \left( f_0 (vi, (v_j, y_j)) - f_0 (v_j) \right)^2$$

(14)

By combining (13) and (14) with (12), we get the desired result.

Now, we discuss the sample error iteratively by applying (11); its necessary is bound to the quantity $\varphi(S, t)$. Hence, we bound for the norm of $f^\lambda_t$ which is required. Suppose that $f^\lambda_{t_0}$ is locally Lipschitz at the origin if the local Lipschitz constant

$$M(\lambda) = \sup \left\{ \frac{l^\prime (u) - l^\prime (0)}{|u|} : |u| \leq \frac{4\kappa^2 |l^\prime (0)|}{\lambda} \right\}$$

is finite for any $\lambda > 0$.

**Lemma 3:** Assume that $l^\prime$ is locally Lipschitz at the origin and $\eta_t (4\kappa^2 M(\lambda) + \lambda) \leq 1$ for each $t$. Then,

$$\left\| f^\lambda_t \right\|_{H_k} \leq 2 \kappa \frac{|l^\prime (0)|}{\lambda}.$$  

(15)

**Proof:** We derive the conclusion by induction. Obviously, $f^\lambda_0 = 0$ satisfies the bound. We assume that this bound holds for $f^\lambda_t$, $\left\| f^\lambda_t \right\|_{H_k} \leq 2 \kappa \frac{|l^\prime (0)|}{\lambda}$. We have

$$\left\| f^\lambda_{t+1} \right\|_{H_k} \leq 2 \kappa \frac{|l^\prime (0)|}{\lambda}.$$  

(16)

By the proof of Lemma 1, we have

$$\left\| f^\lambda_{t+1} \right\|_{H_k}^2 \leq \left\| f^\lambda_t \right\|_{H_k}^2 + \eta_t^2 \left\| f^\lambda_t \right\|_{H_k}^2 + 2\eta_t \varphi(S, t).$$

(17)

By the induction hypothesis, we have

$$\left\| f^\lambda_{t+1} \right\|_{H_k}^2 \leq 2 \kappa \frac{|l^\prime (0)|}{\lambda}.$$  

(18)
\[ f^s_{i+1} = (1 - \eta_i \lambda) f^s_i - \eta_i \left( \frac{m}{2} \right) \]

\[ \sum_{j=1}^{m-1} \sum_{j=m+1}^{m} l'_s(f_i, (v_i, y), (v_j, y)) (K_v - K_y) \]

\[ = (1 - \eta_i \lambda) f^s_i - \eta_i \left( \frac{m}{2} \right) \]

\[ \sum_{j=1}^{m-1} \sum_{j=m+1}^{m} \frac{l'_s(f_i, (v_i, y), (v_j, y)) - l'_s(0)}{f^s_i(v_i) - f^s_i(v_j)} (K_v - K_y) \]

\[ \leq \frac{\eta_i}{m} \left( \frac{m}{2} \right) \]

\[ \sum_{j=1}^{m-1} \sum_{j=m+1}^{m} l'_s(0) (K_v - K_y) \]

Let \( L_y f = \left< f, K_v - K_y \right>_K (K_v - K_y) \). Using

\[ 0 \leq \left< L_y f, f \right>_K = \left| \left< f, K_v - K_y \right>_K \right|^2 \leq 4 \bar{\kappa}^2 \| f \|_{H_K}^2, \]

where \( \| L_y \| \leq 4 \bar{\kappa}^2 \) and \( \frac{l'_s(f_i, (v_i, y), (v_j, y)) - l'_s(0)}{f^s_i(v_i) - f^s_i(v_j)} \) is a positive linear operator on \( H_K \) and its norm is bounded by \( 4 \bar{\kappa}^2 M(\lambda) \).

In terms of \( \eta_i (4 \bar{\kappa}^2 M(\lambda) + \lambda) \leq 1 \), we infer that the operator

\[ A = (1 - \eta_i \lambda) I - \eta_i \left( \frac{m}{2} \right) \]

\[ \sum_{j=1}^{m-1} \sum_{j=m+1}^{m} \frac{l'_s(f_i, (v_i, y), (v_j, y)) - l'_s(0)}{f^s_i(v_i) - f^s_i(v_j)} L_y \]

on \( H_K \) is positive and \( A \leq (1 - \eta_i \lambda) I \). Therefore,

\[ \| f^s_{i+1} \|_{H_K} \leq (1 - \eta_i \lambda) \| f^s_i \|_{H_K} - \frac{\eta_i}{m} \left( \frac{m}{2} \right) \]

\[ \sum_{j=1}^{m-1} \sum_{j=m+1}^{m} \left\| l'_s(0) (K_v - K_y) \right\|_{H_K} = 2 \bar{\kappa} \left\| l'_s(0) \right\| \lambda. \]

For any \( r > 0 \), let \( \zeta = \{ f \in H_K \| \| f \|_{H_K} \leq r \} \), \( L_i = \max\{ \| l'_s(2\bar{\kappa} r) \|, \| l'_s(-2\bar{\kappa} r) \| \} \), and \( M_i = \max\{ \| l(2\bar{\kappa} r) \|, \| l(-2\bar{\kappa} r) \| \} \). For each \( i \in \{1, \ldots, m\} \) and \( v_i' \in V \), we use \( S' \) to denote the sequence obtained from \( S \) by replacing \( v_i \) with \( v_i' \). In the following lemma, we derive the capacity-independent bounds for \( W(S, r) = \sup_{f \in \zeta} \| \hat{R}(f, S) - R(f) \| \).

**Lemma 4:** For every \( r > 0 \) and \( \epsilon > 0 \), we infer

\[ P_S \{ \| W(S, r) - E \{ W(S, r) \} \| < \epsilon \} \]

\[ \leq \exp\left\{ - \frac{2 m^2}{\left( \frac{m}{2} \right)^2} \epsilon^2 \right\} E \{ W(S, r) \} \]

\[ \leq (4L r^2 + 2l(0)) \left( \frac{m}{2} \right)^2. \]

**Proof:** By virtue of (Agarwal & Niyogi, 2009), it is easy to verify that

\[ \| W(S, r) - W(S', r) \| = \sup_{f \in \zeta} \| \hat{R}(f, S) - R(f) \| - \sup_{f \in \zeta} \| \hat{R}(f, S') - R(f) \| \]

\[ \leq \sup_{f \in \zeta} \| \hat{R}(f, S) - R(f, S') \| \leq \frac{m}{2}. \]

Based on MicDiarmid's inequality in (McDiarmid, 1989), we can derive the first result in Lemma 4. Note that

\[ W(S, r) \leq \sup_{f \in \zeta} \| R(f) - \frac{1}{m} \sum_{j=1}^{m} E_i l(f, (v_i, y), (v_j, y)) \|

\[ + \sup_{f \in \zeta} \frac{1}{m} \sum_{j=1, j \neq i}^{m} E_i l(f, (v_i, y), (v_j, y)) - \hat{R}(f, S) \|

\leq \sup_{f \in \zeta} \| E_i l(f, (v_i, y), (v_j, y)) \|

\[ - \frac{1}{m} \sum_{j=1, j \neq i}^{m} l(f, (v_i, y), (v_j, y)) \]

\[ + \frac{1}{m} \sum_{j=1}^{m} \sup_{v_j} E_i l(f, (v_i, y), (v_j, y)) \]

\[ \leq W_1 + W_2. \]
Denote \( G_{\kappa} = \{ h(\nu) = f(\nu) - f(\nu) : \nu \in \zeta \} \). Then,

\[
E\{ W_1 \} = E \sup_{h \in G_{\kappa}} \left( E_{\nu} I(h(\nu)) - \frac{1}{m} \sum_{j=1}^{m} l(h(\nu)) \right)
\]

\[
\leq 2 \sup_{\nu} E \sup_{h \in G_{\kappa}} \left( \frac{1}{m} \sum_{j=1}^{m} \epsilon_j l(h(\nu)) \right)
\]

Since \( l(h(\nu)) - l(0) \leq L_i(h(\nu) - 0) \), we have

\[
E\{ W_1 \} \leq 2L_i \sup_{h \in G_{\kappa}} \left( \frac{1}{m} \sum_{j=1}^{m} \epsilon_j l(h(\nu)) \right) + \frac{2L(0)}{m} \left( E \sum_{j=1}^{m} \epsilon_j \right) \]

\[
\leq \frac{4L_i \kappa r + 2L(0)}{m} \left( E \left( \sum_{j=1}^{m} \epsilon_j \right)^{\frac{1}{2}} \right) \]

\[
= \frac{4L_i \kappa r + 2L(0)}{\sqrt{m}}.
\]

With the same fashion, we can also derive

\[
E\{ W_2 \} \leq \frac{4L_i \kappa r + 2L(0)}{\sqrt{m}}.
\]

By combining (15) and (16), we get the desired conclusion (Liu, Zhu, Jin, & Helali, 2016; Wen, Liu, Wang, & By combining (15) and (16), we get the desired conclusion (Liu, Zhu, Jin, & Helali, 2016; Wen, Liu, Wang, & Caputo, 2016).

Next lemma derives the estimate of \( \varphi(S, t) \).

**Lemma 5:** Assume that \( \eta_i \) satisfies (6) for each \( t \) and \( r = \frac{2L_i \kappa r + 2L(0)}{\sqrt{2L_i \kappa r + 2L(0)}} \). Then, with confidence at least \( 1 - \delta \), we have

\[
\varphi(S, t) \leq B_{\lambda} = (4L_i \kappa r + 2L(0) + M_i \sqrt{2 \log(\frac{2}{\delta})}) \frac{m^{3/2}}{2}.
\]

**Proof:** By virtue of lemma 4, we infer, with confidence at least \( 1 - \delta \),

\[
W(S, r) \leq \frac{4L_i \kappa r + 2L(0)}{\sqrt{m}} \frac{M_i}{2} \sqrt{2 \log(\frac{1}{\delta})}. \text{ Let } f = 0; \text{ in the definition of } f_j, \text{ we get } \frac{1}{2} \left\| f_j \right\|_{L_2} \leq R(0) + 0 \leq \lambda(0).
\]

Then, for any \( \lambda > 0 \), we deduce \( \left\| f_j \right\| \leq \frac{\lambda}{\sqrt{2L_i \kappa r + 2L(0)}} \). Therefore, \( f_j, f_\lambda \in \zeta \), for \( r = \frac{2L_i \kappa r + 2L(0)}{\sqrt{2L_i \kappa r + 2L(0)}} \). Thus, \( \varphi(S, t) \leq 2W(S, r) \) for each \( t \). This finally completes the Lemma 5.

Note that the proposition 3.8 in (Chen, He, & Pan, 2012) is established in our setting.

**Lemma 6:** (Proposition 3.8 in (Chen et al., 2012)) Let \( \eta_i = \eta_i(t^\theta) \) for some \( \theta \in [0, 1] \) and \( \eta_i(4M(\lambda + \lambda) \leq 1 \). Set \( r \) and \( B_1 \) as described in Lemma 5. Let \( B_\lambda = 2\kappa L_i + \kappa \left| I(\nu) \right| \). Then, with confidence at least \( 1 - \delta \), the following bound holds for \( t \geq 1 \): when \( \theta < 1 \),

\[
\left\| f_j \right\|_{H_\lambda} \leq \left\| f_j \right\|_{H_\lambda} \exp \left( -\frac{\lambda \eta_i}{1 - \theta} (t^{1-\theta} - 1) \right) + \frac{18B_\lambda^2}{\lambda \eta_i} t^{1-\theta} + \frac{6B_\lambda}{\lambda} - 1 \theta \exp \left( 1 - \theta \right) (t + 1) + \frac{6B_\lambda}{\lambda}.
\]

**4. Proof of main result**

According to the lemmas we derived in Section 3, we now present the detailed proof of Theorem 1. Let \( C_1 = 2\kappa \left| I(\nu) \right| + 2L(0) \). By \( \lambda = t^\theta \), we have \( r \leq C_1 \epsilon^\theta \). Using (4), we get \( L_i \leq C_0(1 + 2\kappa \epsilon^{1-\theta}) \leq C_0(1 + 2\kappa \epsilon^{1-\theta} t^{1-\theta}) \) and \( M_i \leq C_0(1 + 2\kappa \epsilon^{1-\theta}) \). Let \( C_4 = 4C_0(1 + 2\kappa \epsilon^{1-\theta}) \) and \( C_5 = 2C_0(1 + 2\kappa \epsilon^{1-\theta}) + 2\phi(0) \). We infer \( B_\lambda \leq C_4 \sqrt{2 \log(\frac{2}{\delta})} \frac{m^{3/2}}{2} \). If \( 1 \leq \left| t \right| \leq \frac{4\kappa \epsilon^{1-\theta}}{\lambda} \), we obtain \( \left| I(\nu) - I(\nu) \right| \leq \left| I(\nu) \right| + \frac{2\kappa \epsilon^{1-\theta}}{\lambda} \). Therefore, \( M(\lambda) \leq \sup_{|t| \leq \frac{4\kappa \epsilon^{1-\theta}}{\lambda}} \left| I(\nu) \right| + \frac{18B_\lambda^2}{\lambda \eta_i(t^{\theta+1} - 1 + \frac{6B_\lambda}{\lambda})}. \)

Thus, the condition \( \eta_i(4M(\lambda + \lambda) \leq 1 \) in Lemma 6 establishes if \( \eta_i = \eta_i(t^\theta) \) and \( \eta_i \eta_i(\text{max}(q(1-\theta)^v) \). In view of Lemma 6 and \( \left( \frac{m^{3/2}}{2} \right)^2 \leq \left( \frac{m^{3/2}}{2} \right)^2 \), we yield, with confidence at least \( 1 - \delta \),

\[
\left\| f_j \right\|_{H_\lambda} \leq \left( \frac{\lambda \eta_i}{1 - \theta} (t^{1-\theta} - 1) \right) \left( \frac{m}{2} \right)^{3/2} \left( \frac{2}{m} \right)^{1-\theta} \left( \frac{1}{2} \right)^{(1-\theta)(2(1-\theta))}.
\]
\[ \exp\left(-\frac{\eta(1-\tilde{c}\tilde{r})}{1-\theta} + \theta \max(q-1,1) \right) + \tilde{c}_4 \left( \frac{m}{m^2} \right)^{\min(q+1,2q-1)-\theta} \]

where \( \tilde{c}_p, \tilde{c}_2, \tilde{c}_3, \) and \( \tilde{c}_4 \) are constants independent on \( m \) and \( t \). Thus, if \( 1 - \theta - y \max\{q - 1, 1\} > 0 \), we finally get the desired result.

Acknowledgments

First, we thank the reviewers for their constructive comments in improving the quality of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Agarwal, S., & Niyogi, P. (2009). Generalization bounds for ranking algorithms via algorithmic stability. *Journal of Machine Learning Research*, 10, 441–474.

Bouzeghoub, A., & Elbyed, A. (2006). Ontology mapping for web-based educational systems interoperability. *Interoperability in Business Information Systems*, 1, 73–84.

Chen, H., He, F., & Pan, Z. (2012). Approximation analysis of gradient descent algorithm for bipartite ranking. *Journal of Applied Sciences*, 39, 151–159.

Fonseca, F., Egenhofer, E., Davis, C., & Camara, G. (2001). Semantic granularity in ontology-driven geographic information systems. *AMAI Annals of Mathematics and Artificial Intelligence – Special Issue on Spatial and Temporal Granularity*, 36, 121–151.

Gao, W., & Liang, L. (2011). Ontology similarity measure by optimizing NDCG measure and application in physics education, future communication, computing. *Control and Management*, 142, 415–421.

Gao, W., & Shi, L. (2013). Ontology similarity measure algorithm with operational cost and application in biology science. *BioTechnology: An Indian Journal*, 8, 1572–1577.

Gao, W., & Xu, T. (2012). Characteristics of optimal function for ontology similarity measure via multi-dividing. *Journal of Networks*, 7, 1251–1259.

Gao, W., & Xu, T. (2013). Stability analysis of learning algorithms for ontology similarity computation. *Abstract and Applied Analysis*, 2013, 1–9.

Gao, W., Gao, Y., & Liang, L. (2013). Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping. *Journal of Chemical and Pharmaceutical Research*, 5, 592–598.

Gao, W., Gao, Y., & Zhang, Y. (2012). Strong and weak stability of k-partite ranking algorithm. *Information*, 15, 4585–4590.

Gao, W., Gao, Y., Zhang, Y., & Liang, L. (2014). Minimax learning rate for multi-dividing ontology algorithm. *Journal of Information and Computational Science*, 11, 1853–1860.

Gao, W., Xu, T., Gan, J., & Zhou, J. (2014). Linear statistical analysis of multi-dividing ontology algorithm. *Journal of Information and Computational Science*, 11, 151–159.

Gao, W., Yan, L., & Liang, L. (2014). Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (I). *International Journal of Computer Applications in Technology*, 50, 226–231.

Gao, Y., & Gao, W. (2012). Ontology similarity measure and ontology mapping via learning optimization similarity function. *International Journal of Machine Learning and Computing*, 2, 107–112.

Gao, Y., Gao, W., & Liang, L. (2016). Statistical characteristics for multi-dividing ontology algorithm in AUC criterion setting. *International Journal of Collaborative Intelligence*, 3, 178–188.

Hu, B., Dasmahapatra, S., Lewis, P., & Shadbolt, N. (2003). Ontology-based medical image annotation with description logics. *15th IEEE International Conference on Tools with Artificial Intelligence*, Washington, DC, pp. 77–82.

Huang, X., Xu, T., Gao, W., & Gong, S. (2011). Ontology similarity measure and ontology mapping using half transductive ranking. *Proceedings of 2011 4th IEEE International Conference on Computer Science and Information Technology*, Chengdu, pp. 571–574.

Huang, X., Xu, T., Gao, W., & Jia, Z. (2011). Ontology similarity measure and ontology mapping via fast ranking method. *International Journal of Applied Physics and Mathematics*, 1, 54–59.

Lan, M., Xu, J., & Gao, W. (2012). Ontology similarity computation using k-partite ranking method. *Journal of Computer Applications*, 32, 1094–1096.

Lin, Z., Weige, T., Xiaozhong, M., & Wei, G. (2016). Theoretical characteristics of ontology learning algorithm in multi-dividing setting. *IAENG International Journal of Computer Science*, 43, 184–191.

Liu, W., Zhu, L. Y., Jin, B. B., & Helali, M. (2016). Dynamic modeling of parallel mechanism based on particle system. *Journal of Mechanical Engineering Research and Developments*, 39, 483–491.

McDiarmid, C. (1989). On the method of bounded differences. *In Surveys in Combinatorics, 141*, 148–188.

Wang, Y., Gao, W., Zhang, Y., & Gao, Y. (2010). Ontology similarity computation use ranking learning method. *The 3rd International Conference on Computational Intelligence and Industrial Application*, Wuhan, pp. 20–22.

Wen, H. X., Liu, X. J., Wang, H., & Caputo, Y. (2016). The influence of financial constraint and agency cost on investment inefficiency of GEM listed firms-based on two-tier stochastic frontier model. *Journal of Mechanical Engineering Research and Developments*, 39, 584–592.

Yan, L., Gao, W., & Li, J. (2013). Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (II). *Journal of Applied Sciences*, 13, 3257–3262.

Yang, Y., & Zhou, D.X. (2006). Online regularized classification algorithms. *IEEE Transactions on Information Theory*, 52, 4775–4788.