Adaptation Procedure for Multi-dividing Ontology Algorithm

Wei Gao, Yun Gao and Li Liang
1School of Information Science and Technology,
Yunnan Normal University, 650500, Kunming, China
2Editorial Department of Yunnan Normal University, 650092, Kunming China

Abstract: Ontology, as a structured conceptual model of knowledge representation and storage, has widely used in computer science, biomedical and social science. Ontology algorithms even become the core algorithms in information retrieval and thus raise more and more attention. In this study, we propose the adaptation procedure for ontology algorithm in multi-dividing setting. The new technology helps to adapt an unknown parameter and we determine the generalization error bound for such adaptation procedure under certain assumption.

Key words: Ontology, similarity measure, ontology mapping, multi-dividing, adaptation procedure

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 social science (Bouzeghoub and Elbayed, 2006), biology medicine (Hu et al., 2003) and geography science (Ponsese et al., 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. Let graphs G1, G2, ..., Gk corresponding to ontologies O1, O2, ..., Ok respectively and G = G1+G2+...+Gk. For every vertex v∈V(Gi), where 1≤i≤k, the goal of ontology mapping is finding similarity vertices from G-Gi. Thus, the core of ontology mapping problem is just ontology similarity measure.

There are some effective technologies for ontology similarity measure. Wang et al. (2010) first proposed that ranking method can be employed in ontology similarity calculation. Huang et al. (2011a) raised fast ontology algorithm in order to reduce the time complexity of the algorithm. Gao and Liang (2012) argued that ontology function can 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. In Huang et al. (2011b) the proposal was getting ontology function based on half transductive ranking. Lan et al., 2012 explored the learning theory approach for ontology similarity computation in a setting when the ontology graph is a tree. Using the multi-dividing algorithm in which the vertices divided into k parts correspond to the k classes of rates. The rate values of all classes are decided by experts. Then, a vertex in a rate a has larger value than any vertex in rate b (where, 1≤a<b≤k) under ontology function f. Finally, the similarity between two ontology vertices is measured by the difference of two real numbers which they correspond to. Thus, the multi-dividing algorithm is reasonable to learn a score function for an ontology graph with a tree structure. Gao et al. (2013a) presented new algorithms for ontology similarity measurement and ontology mapping using 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.


Corresponding Author: Wei Gao, School of Information Science and Technology, Yunnan Normal University, 650500, Kunming, China

3005
(2014a) 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 et al. (2013b) and Yan et al. (2013) presented an approach of piecewise constant function approximation for AUC criterion multi-dividing ontology algorithms.

In this study, we focus on the adaptation procedure of multi-dividing ontology algorithm. The contributions of this study are two-fold: First, adaptation procedure for multi-dividing ontology setting is proposed and second, we determine the statistic error bound for such adaptation procedure under certain assumption. The organization of this study is as follows: The multi-dividing ontology problem and notations for basic setting are showed in Section 2, the adaptation procedure model for such setting stated in Section 3 and also we present some preliminary terminology and lemma; in Section 4, we determine the main conclusion in this study. We present the statistic error bound for adaptation procedure method, in last Section, we discuss some extension problem for further researching.

SETTING

Using a vector with dimension q to express all the information for each vertex in ontology graph. Then, we say V takes its value in a high dimension feature space $V \in \mathbb{R}^q$. The elements in V are drawn independently and randomly according to some unknown distribution D. Given a training sample set $S = \{v_1, ..., v_n\} \subset V$ 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 there k classes.

Formally, the settings of 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 $S = (S_1, S_2, ..., S_k) \in V^{n \times V^{n \times ..., V^{n}}}$ consisting of a sequence of training sample $S_i = (v_1, ..., v_n)$ (1 $\leq$ $i$ $\leq$ k) with $|S_i| = n_i$. The goal is to learn from these samples a real-valued ontology score function $f: V \rightarrow \mathbb{R}$ that orders the future S vertices rank higher than $S_i$ where $a$<$b$. We assume that instances in each $S_i$ are drawn independently and according to some (unknown) distribution $D_i$ on the instance space $V$ respectively. Denote Y be a label of $V$ which indicate its classification information.

For $v, v' \in V$, the order of $x$ and $x'$ under multi-dividing ontology function $f$ should be consistent with their labels. The ontology loss function $l: \mathbb{R}^{V} \times V \times V' \rightarrow \{0, 1\}$ is used to punish inconsistency situation which $\text{sgn}(f(v') - f(v))$ is not coincide with their relationships, where:

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

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

Let $F$ be a class of measurable functions form $V$ to $\mathbb{R}$. The quality of multi-dividing ontology function is measured by the expected l-error:

$$R_l(f) = \frac{1}{|S|} \sum_{i=1}^{n} \sum_{j=1}^{n_i} b_{i} a_{ij} \sum_{b=1}^{n} \sum_{c=1}^{n_b} a_{ij} b_{bc} l(f(v, v'))$$

But it cannot be estimated directly since distribution D is unknown. We use empirical l-error to measure multi-dividing ontology algorithm instead:

$$\hat{R}_l(f; S) = \frac{1}{|S|} \sum_{i=1}^{n} \sum_{j=1}^{n_i} b_{i} a_{ij} \sum_{b=1}^{n} \sum_{c=1}^{n_b} a_{ij} b_{bc} l(f(v', v'))$$

Multi-dividing ontology algorithm that, given a training sample $S$, output a multi-dividing ontology function $F_{\hat{f}} \in F$ that satisfies:

$$f_{\hat{f}} = \arg\min_{f \in F} \hat{R}_l(f; S)$$

(1)

ADAPTATION PROCEDURE MODEL AND USEFUL LEMMA

In this section, we construct cross-validation adaptation procedures for multi-dividing ontology setting. A sequence of functions $\hat{f} = (f^{(r)})_{r=1}^{\infty}$ is called a statistic if each map $f^{(r)}$ associating a function $f^{(r)}(\cdot) = f^{(r)}(S(\cdot))$ in $F$ to each data set $S = (S_1, S_2, ..., S_k)$. The risk of its n-th element $\hat{f}^{(n)}$ in multi-dividing setting is denoted as:

$$R(\hat{f}^{(r)}(S)) = E[|\hat{f}(v) - V(v)|]$$

(2)

We define the multi-dividing ontology adaptation procedures empirical risk for a statistic $\hat{f} = (f^{(r)})_{r=1}^{\infty}$ by:

$$R_{\hat{f}}(f) = \frac{1}{|S|} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{b=1}^{n} \sum_{c=1}^{n_b} |l(\hat{f}(v, v'))|$$

(3)
The multi-dividing ontology adaptation procedure for fixed $p$ statistics $\tilde{f}_1, \ldots, \tilde{f}_p$ is the adaptation procedure $\tilde{f}_a = (\tilde{f}_a^0)_{a=1}^m$ defined by:

$$\tilde{f}_a^0(S) = \tilde{f}_a(S)$$

with:

$$j(S) \in \arg \min_{j(S)} R_{\kappa}(f)$$

A Young function $\psi: \mathbb{R}^+ \rightarrow \mathbb{R}$ (Van Der Vaart and Wellner, 1996) is non-decreasing and convex function with $\psi(0) = 0$ and $\psi(\infty) = \infty$. The $\psi$-norm of multi-dividing ontology function $f$ is given by:

$$\|f\|_\psi = \inf \{C > 0 : E\psi(\frac{|f|}{C}) \leq 1\}$$

Let $\psi = \exp(x^\alpha)-1$ for $\alpha \geq 1$.

Now, we propose following assumption for the excess loss function of an estimator $\hat{f}$ in multi-dividing ontology setting.

**Assumption 1**: For each sample set $S = (S_1, S_2, \ldots, S_k)$, there exist $\kappa \geq 1$, $K_0$ and $K_\psi > 0$ satisfies:

$$\|f(S) - f^*(S)\|_{\psi} \leq K_0$$

$$\|f(S) - f^*(S)\|_{\psi} \leq K_\psi (R(f(S)) - R(f^*))^{1/\alpha}$$

For given sample size, we require certain characters on the estimators $\tilde{f}_1, \ldots, \tilde{f}_p$ to achieve a statistic inequality for the adaptation procedure in multi-dividing ontology setting. A statistic $f = (\tilde{f}_a^0), a=1, \ldots, n$, is called exchangeable if for any fixed integer $n_1, \ldots, n_k$ and any permutation $\phi: \{1, \ldots, n\} \rightarrow \{1, \ldots, n\}$ for each $a \in \{1, \ldots, k\}$, we obtain $\tilde{f}_a^0(S) = \tilde{f}_a^0(\kappa(S))$.

The main results in our study depend heavily on the following lemma which determined by Lecue and Mitchell (2012):

**Lemma 1**: Let $L = \{L_1, \ldots, L_n\}$ be a set of $p \geq 1$ measurable functions defined on $(V, \Gamma)$. Let $V, V_1, \ldots, V_n$ be i.i.d. random variables with values in $(V, \Gamma)$ such that $\forall L_a, E(L_a) = 0$. Assume the existence of constants $C_p > 0$ and $\kappa \geq 1$ such that $\forall L_a, L(V)|L_a| \leq C_p (EQ(V))^{1/2}$. Let:

$$b_n = \max_a \max_L \max |L_a| \leq 1$$

For any given shift parameter $\epsilon > 0$ there exists a constant $c^* = c^*(\epsilon, \kappa)$ such that:

$$\max_a \max_L \max |L_a| \leq \frac{1}{m} \sum_{i=1}^m L(V_i)$$

$$\leq c^* \left( \frac{\log(ep)}{m} \right)^{1/\alpha} \left( \frac{1}{m} \sum_{i=1}^m \log(ep) \right)$$

**MAIN RESULT AND PROOF**

Beginning on the way to show and proof our main result, we first present following lemma which determines that supremum bounds on the adaptation empirical process for the learned estimates $\tilde{f}_a^0(S)$ in multi-dividing ontology setting.

**Lemma 2**: Suppose that the ontology risk map $\mathbf{R}(\mathbf{r})$ is convex and estimator $\tilde{f}_a^0 - \tilde{f}_a^0(S)$ is the version of the adaptation procedure (4). Let:

$$P_S = \sum_{a=1}^m \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \sum_{k=1}^{n_c} \delta_{a,i,j,k}$$

be the empirical probability measure. Then for each constant $\epsilon > 0$, we have:

$$\mathbb{E}_p \left( \mathbf{R}(\tilde{f}_a^0(S)) - \mathbf{R}(f^*) \right)$$

$$\leq (1 + \epsilon) \min_{\alpha(n)} [\mathbb{E}_p \left( \mathbf{R}(\tilde{f}_a^0(S)) - \mathbf{R}(f^*) \right) + \mathbb{E}_p \max_{i,j,k} \left[ (1 + \epsilon) P_S \mathbf{R}(\tilde{f}_a^0(S)) \right)]$$

**Proof**: First, we consider the conclusion for $\tilde{f}_a^0 = \tilde{f}_a^0$. In terms of $j(S)$, we infer, for each $j \in \{1, \ldots, p\}$:

$$\mathbb{R}_{\kappa}(f(S)) = \sum_{a=1}^m \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \sum_{k=1}^{n_c} \delta_{a,i,j,k} \mathbb{R}_{\kappa}(f(S), V^i, V_j) \leq \mathbb{R}_{\kappa}(f(S))$$

By virtue of Eq. 5, we yield the following fact for any $j$ and each data set $S$:

$$\mathbb{R}(\tilde{f}_a^0(S)) - \mathbb{R}(f^*) = (1 + \epsilon) \mathbb{R}_{\kappa}(f(S)) + \mathbb{R}(\tilde{f}_a^0(S)) - \mathbb{R}(f^*) (1 + \epsilon) \mathbb{R}_{\kappa}(f(S))$$

$$\mathbb{R}(\tilde{f}_a^0(S)) - \mathbb{R}(f^*) = (1 + \epsilon) \mathbb{R}_{\kappa}(f(S)) + \mathbb{R}(\tilde{f}_a^0(S)) - \mathbb{R}(f^*)$$

$$
\mathbb{R}(\tilde{f}_a^0(S)) - \mathbb{R}(f^*) = (1 + \epsilon) \mathbb{R}_{\kappa}(f(S)) + \mathbb{R}(\tilde{f}_a^0(S)) - \mathbb{R}(f^*)
$$

Since the $V_j$'s are i.i.d., it follows that the expectation of the first term in Eq. 6 is such that for every $j$:

$$\mathbb{E}_p \mathbb{R}_{\kappa}(f(S)) - \mathbb{E}_p \mathbb{R}_{\kappa}(f^*)$$

$$= \sum_{i,j,k} \frac{1}{m} \sum_{a=1}^m \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \sum_{k=1}^{n_c} \delta_{a,i,j,k} (\mathbb{E}_p \mathbb{R}(\tilde{f}_a^0(S), V^i, V_j) - \mathbb{E}_p \mathbb{R}(f^*, V^i, V_j))$$

$$= \sum_{i,j,k} \frac{1}{m} \sum_{a=1}^m \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \sum_{k=1}^{n_c} \delta_{a,i,j,k} (\mathbb{E}_p \mathbb{R}(\tilde{f}_a^0(S)) - \mathbb{R}(f^*)) = \mathbb{E}_p \mathbb{R}(\tilde{f}_a^0(S)) - \mathbb{R}(f^*)$$
In view of the convexity of the risk, we deduce:

\[ E_x[R(\hat{f}_w(S)) - R(f^*) - (1 + \epsilon)\hat{R}_h(\hat{f}_w) - \hat{R}_h(f^*)] \]

\[ \leq \sum_{i=1}^{k} \sum_{l=1}^{k} \sum_{s=1}^{n} \sum_{v=1}^{n} \sum_{v' \in V} [P(f^{(s)}_w(S), v, v') - P(f^{(s)}_w(S), v', v')]

\[ - \frac{1 + \epsilon}{n \cdot n} \sum_{i=1}^{k} \sum_{l=1}^{k} \sum_{s=1}^{n} \sum_{v=1}^{n} \sum_{v' \in V} (l(f^{(s)}_w(S), v, v') - l(f^{(s)}_w(S), v', v')) \]

\[ \leq \sum_{i=1}^{k} \sum_{l=1}^{k} E_x \max_{s \in V, v \in V} \sum_{s=1}^{n} \sum_{v=1}^{n} \sum_{v' \in V} [P(f^{(s)}_w(S), v, v') - P(f^{(s)}_w(S), v', v')]

\[ - (l(f^*, v, v') - l(f^*, v', v')) = E_x \max_{s \in V, v \in V} \sum_{s=1}^{n} \sum_{v=1}^{n} \sum_{v' \in V} (l(f^{(s)}_w(S), v, v') - l(f^{(s)}_w(S), v', v')) \]

Thus, we complete the proof of Lemma 2.

By combining Lemma 1 and 2, we get the following statistic inequality for the adaptation procedures in multi-dividing ontology setting. Proven method is followed by the classical techniques of statistical learning theory. We omit the detail proof.

Theorem 1: Let \( \tilde{f}_1, \ldots, \tilde{f}_p \) be p statistics satisfying Assumption 1. Suppose that the risk map \( f^* - R(f) \) is convex and estimator \( \tilde{f}_w^* - \tilde{f}_w \) is the version of the adaptation procedure (4). Then for each positive constant \( \epsilon \), there exist a constant \( c' = c'(\epsilon, \kappa) \) such that:

\[ E_x[R(\tilde{f}_w^*(S)) - R(f^*)] \leq (1 + \epsilon) \min_{\rho \in \rho^*} E_x[R(\tilde{f}_w(S)) - R(f^*)] \]

\[ + \frac{1}{c'} \log \left( \frac{1}{n \cdot n} \sum_{s=1}^{n} \sum_{v=1}^{n} (\log \log n \cdot \log \frac{p}{n}) \right) \]

DISCUSSION

Let \( \Lambda \) be a set of indexes and \( F = (\tilde{f}_\lambda, \lambda \in \Lambda) \) be a set of statistics generated by \( \Lambda \). In the previous part of our article we discuss the case \( \Lambda = \{1, \ldots, p\} \). For more generally setting, we need not suppose \( \Lambda \) to be finite. We define the continuous version of the multi-dividing ontology adaptation procedure by:

\[ \tilde{f}_w(S) = \tilde{f}_w^{(s)}(S) \]

(7)

Where:

\[ \hat{\lambda}(S) = \arg \min_{\lambda \in \Lambda} \tilde{f}_w^{(s)}(S) \]

Hence, the statistic characters for adaptation process in multi-dividing ontology setting such that any statistic \( \tilde{f}_w \) in \( F \) satisfies Assumption 1 and estimator \( \tilde{f}_w^* = \tilde{f}_w^{(s)} \) is the version of the adaptation procedure 7 will be consider in the furture.

REFERENCES


