Chaos, Solitons and Fractals 89 (2016) 322–327

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Leave-two-out stability of ontology learning algorithm☆

Jianzhang Wu[a,*], Xiao Yu[b], Linli Zhu[c], Wei Gao[d]

[a] School of Computer Science and Engineer, Southeast University, Nanjing 210096, China
[b] School of Continuing Education, Southeast University, Nanjing 210096, China
[c] School of Computer Engineering, Jiangsu University of Technology, Changzhou, Jiangsu 213001, China
[d] School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China

ARTICLE INFO

Article history:
Available online 7 January 2016

Keywords:
Ontology
Similarity measure
Learning algorithm
Ontology loss function
Stability

ABSTRACT

Ontology is a semantic analysis and calculation model, which has been applied to many subjects. Ontology similarity calculation and ontology mapping are employed as machine learning approaches. The purpose of this paper is to study the leave-two-out stability of ontology learning algorithm. Several leave-two-out stabilities are defined in ontology learning setting and the relationship among these stabilities are presented. Furthermore, the results manifested reveal that leave-two-out stability is a sufficient and necessary condition for ontology learning algorithm.

1. Introduction

It is in philosophy that the term "ontology" is first applied to describe the connection nature of things and the inherently hidden connections of their components. Ontology, being a model for storing and representing knowledge, has been widely applied in knowledge management, machine learning, information systems, image retrieval, information retrieval search extension, collaboration and intelligent information integration in information and computer science. Meanwhile, ontology is an effective concept semantic model and a powerful analysis tool. It has been used extensively in pharmacology, biology, medicine, geographic information system and social science in the past ten years (see Przydzial et al., [1], Koehler et al., [2], Ivanovic and Budimac [3], Hristoskova et al., [4], and Kabir [5]).

A simple graph can be used to express the structure of ontology. On that graph, each vertex represents a concept, object or element in ontology. Each (directed or undirected) edge refers to a relationship or hidden connection between two concepts (objects or elements). Let O be an ontology and G be a simple graph of O. The purpose of ontology engineer application is to get the similarity calculating function and then compute the similarities between ontology vertices. The inherent association between vertices in ontology graph can be illustrated by these similarities. Ontology mapping is to obtain the ontology similarity measuring function by measuring the similarity between vertices from different ontologies. Such a mapping connects different ontologies, through which a potential link between the objects or elements from different ontologies can be acquired. The ontology similarity function $Sim: V \times V \to \mathbb{R}^+ \cup \{0\}$ is a semi-positive score function mapping in which each pair of vertices maps to a non-negative real number.

An advanced usage of handling the ontology similarity computation is using ontology learning algorithm which gets an ontology function $f: V \to \mathbb{R}$. Using such an ontology function, the ontology graph is mapped into a line consisting of real numbers. After comparing the difference between their corresponding real numbers, the similarity

☆ This work was supported in part by the National Natural Science Foundation of China (60903131), Science and Technology of Jiangsu Province (BE2011173), and Key Laboratory of Computer Network and Information Integration Founding in Southeast University.

* Corresponding author. Tel.: +86 4403877466372828.
E-mail address: jzwu@njnet.edu.cn (J. Wu).

http://dx.doi.org/10.1016/j.chaos.2015.12.013
0960-0779/© 2015 Elsevier Ltd. All rights reserved.

between two concepts can be measured and in which the dimensionality reduction is the ore of the idea. If the ontology function is to be associated with ontology application, a vector is a very good choice as it expresses all the information of a vertex, for instance v. In a simpler representation, the notations are slightly confused and v is used to denote both the ontology vertex and its corresponding vector. The vector is mapped to a real number by ontology function $f : V \to \mathbb{R}$. The ontology function, which is a dimensionality reduction operator, maps vectors of multi-dimension into one dimensional ones.

All the related information of an arbitrary vertex in ontology graph G is expressed by a p dimensional vector, which includes its instance, structure, name, attribute, and semantic information of the concept which is corresponding to the vertex and that is contained in its vector. In order not to lose generality, it can be assumed that $v = \{v_1, \ldots, v_p\}$ is a vector corresponding to a vertex v. Their notations are slightly confused and v is adopted to represent both the ontology vertex and its corresponding vector. In order to obtain an optimal ontology function $f : V \to \mathbb{R}$, ontology learning algorithms are used by the authors. Therefore, the value of $|f(v_i) - f(v_j)|$ is used to determine the similarity between two vertices v_i and v_j. Dimensionality reduction, i.e., using real number to represent p dimension vector is the core of such kind of ontology algorithm. In this way, we can regard an ontology function f as a dimensionality reduction operator $f : \mathbb{R}^p \to \mathbb{R}$.

There are many effective methods for getting efficient ontology similarity measure or ontology mapping algorithm. They have been studied in terms of ontology function. Moreover, the theoretical research of ontology algorithms has been contributed by several researchers. The uniform stability of multi-dividing ontology algorithm and the generalization bounds for stable multi-dividing ontology algorithms was put forth by Gao and Xu [6]. A gradient learning model for ontology similarity measuring and ontology mapping in multi-dividing setting was proposed by Gao and Zhu [7] cooperatively. In the setting, the sample error was determined in terms of the hypothesis space and the ontology dividing operator in which one can suppose that V is an instance space.

In this article, we research the influences of ontology learning algorithm when two sample vertices are deleted from ontology sample set. In next section, we describe the detailed notations, definitions and setting of ontology learning problem. And then, the main conclusions are drawn in Section 3.

2. The notations, definitions and setting of ontology learning problem

Suppose that V is a compact domain in Euclidean space and Y is a set of labels. Let $\mu(v, y)$ be an unknown probability distribution on $Z = V \times Y$ and $S = \{(v_1, y_1), \ldots, (v_n, y_n)\} = (v_i, y_i)_{i=1}^n = (z_i)_{i=1}^n$ be an ontology sample set consisting of n samples drawn i.i.d. from the probability distribution on Z_n. The aim of ontology learning is to predict an ontology function $f_S : V \to \mathbb{R}$ using the empirical ontology data S which evaluates at a new ontology vertex v to predict its corresponding value of y.

Let $L : \mathbb{R}^V \times V \times \mathbb{R} \to \mathbb{R}$ be the ontology loss function and $L(f, z)$ be the value of punishment for fixed ontology function f and $z = (v, y)$. Throughout our paper, we always assume that the loss function L is square ontology loss $L(f, z) = (f(v) - y)^2$ and there exists M which satisfies $0 \leq L(f, z) \leq M$ for any $f \in \mathcal{F}$ (here $\mathcal{F}$ is a hypothesis space in ontology setting) and $z \in Z$. Denote $l(z) = L(f, z)$ for convenience. Thus $l(z) : V \times Y \to \mathbb{R}$ and we set $\mathcal{L} = \{l(f) : f \in \mathcal{F}\}$ as the space of ontology loss function.

The ontology expected error for fixed ontology function f, ontology loss function L and a probability distribution μ is defined by: $R(f) = \mathbb{E}_Z L(f, z)$. When L is square ontology loss, we have

$$R(f) = \mathbb{E}_Z L(f, z) = \int_{V,Y} (f(v) - y)^2 d\mu(v, y) = \mathbb{E}_\mu |f(v) - y|^2.$$

However, $R(f)$ can't be calculated directly since μ is unknown. In reality, we compute the ontology empirical error instead which is presented as $\widehat{R}_S(f) = \frac{1}{n}\sum_{i=1}^n L(f, z_i)$. In addition, in our square ontology loss setting, it is equal to $\widehat{R}_S(f) = \frac{1}{n}\sum_{i=1}^n (f(v_i) - y_i)^2 = \mathbb{E}_{\mu_n}(f(v) - y)^2$, where μ_n is the ontology empirical supported on $\{v_1, \ldots, v_n\}$ which means $\mu_n = \sum_{i=1}^n \delta_{v_i}/n$ and δ_{v_i} is the vertex evaluation functional on v_i.

In what follows, set $S^{i,j}$ as the ontology training set from S by deleting two vertices v_i and v_j $(1 \leq i < j \leq n)$. For our ontology learning setting, the functions f_S and $f_{S^{i,j}}$ are the minimizers of $\widehat{R}_S(f)$ and $\widehat{R}_{S^{i,j}}(f)$, respectively. The notations $\mathbb{E}_S$ and $\mathbb{P}_S$ are used to express the expectation and the probability on the ontology training set S which is drawn i.i.d, according to probability distribution on Z_n.

An ontology algorithm is called symmetric if the optimal ontology function can't be changed when the elements in training set S are re-arranged. Given an ontology training set S and an ontology function space $\mathcal{F}$, the almost ontology learning algorithm is defined as a symmetric procedure which chooses an ontology function $f_S^{\varepsilon^E}$ that minimizes the ontology empirical risk over all ontology functions $f \in \mathcal{F}$, we infer

$$\widehat{R}_S(f_S^{\varepsilon^E}) \leq \inf_{f \in \mathcal{F}} \widehat{R}_S(f) + \varepsilon^E, \tag{1}$$

where $\varepsilon^E > 0$.

An ontology learning map is called (universally, weakly) consistent if for any positive number $\varepsilon_c > 0$, we have

$$\lim_{n \to \infty} \sup_\mu \mathbb{P}\{R(f_S) > \inf_{f \in \mathcal{F}} R(f) + \varepsilon_c\} = 0.$$

The consistency is universal implies that the inequality above is established with respect to the set of any measure on Z, whereas weak consistency means only convergence in probability and strong consistency requires almost sure convergence.

Let $\mathcal{F}$ be any class of functions. $\mathcal{F}$ is a (weak) uniform Glivenko–Cantelli (in short, uGC) class if for any positive number ε,

$$\lim_{n \to \infty} \sup_\mu \mathbb{P}\{\sup_{f \in \mathcal{F}} |\mathbb{E}_{\mu_n} f - \mathbb{E}_\mu f| > \varepsilon\} = 0.$$

When comes to the ontology loss functions l, the definition implies that for all distributions μ there exist ε_n and $\delta_{\varepsilon_n, n}$

such that $\lim_{n\to\infty}\varepsilon_n=0$, $\lim_{n\to\infty}\delta_n=0$ and

$$\mathbb{P}\{\sup_{f\in\mathcal{F}}|R[(f)-\widehat{R}_S(f)|>\varepsilon_n\}\le\delta_{\varepsilon_n,n}.$$

For any ontology function space $\mathcal{F}$, it is called a strong uniform Glivenko–Cantelli class if for any positive number ε, we deduce

$$\lim_{n\to\infty}\sup_{\mu}\mathbb{P}\{\sup_{m\ge n}\sup_{f\in\mathcal{F}}|\mathbb{E}_{\mu_m}f-\mathbb{E}_{\mu}f|>\varepsilon\}=0.$$

For bounded ontology loss functions, weak uGC is equivalent to strong uGC and the weak consistency is equivalent to strong consistency.

The uniform stability of ontology learning algorithm is stated as for any $S\in Z^n$, $i,j\in\{1,\dots,n\}$, we have:

$$\sup_{z\in Z}|L(f_S,z)-L(f_{S^{i,j}},z)|\le\beta. \tag{2}$$

The $(\beta,\ \delta)$ hypothesis stability of ontology learning algorithm is stated as follows which is a natural criterion for hypothesis spaces (here $\lim_{n\to\infty}\beta=0$ and $\lim_{n\to\infty}\delta=0$),

$$\mathbb{P}_S\{\sup_{z\in Z}|L(f_S,z)-L(f_{S^{i,j}},z)|\le\beta\}\ge1-\delta. \tag{3}$$

Similarly, the cross-validation in leave-two-out setting ($CV_{\rm lto}$) stability of ontology learning algorithm can be presented that for any $k\in\{1,\dots,n\}$, we infer

$$\mathbb{P}_S\{|L(f_S,z_k)-L(f_{S^{i,j}},z_k)|\le\beta_{CV}\}\ge1-\delta_{CV}.$$

In fact, uniform stability reveals $(\beta,\ \delta)$ hypothesis stability which reveals $CV_{\rm lto}$ stability in ontology learning setting.

The following facts are easily to check and will be used in the next section.

$$\mathbb{E}_S[R(f_S)]=\mathbb{E}_S[\mathbb{E}_zV(f_S,z)]=\mathbb{E}_{S,z}[V(f_S,z)],$$

for all $i\in\{1,\dots,n\}$, and

$$\mathbb{E}_S[R(f_S)]=\mathbb{E}_S\left[\frac{1}{n}\sum_{i=1}^{n}L(f_S,z_i)\right]=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_S[L(f_S,z_i)]$$
$$=\mathbb{E}_S[V(f_S,z_i)].$$

The ontology learning map A is called distribution-independent, $(\beta_{CV}^{(n)},\delta_{CV}^{(n)})$ $CV_{\rm lto}$ stable if for any positive integer n there exist a $\beta_{CV}^{(n)}$ and $\delta_{CV}^{(n)}$ such that for any $k\in\{1,\dots,n\}$ and any distribution μ it satisfies $\lim_{n\to\infty}\beta_{CV}^{(n)}=0$, $\lim_{n\to\infty}\delta_{CV}^{(n)}=0$ and

$$\mathbb{P}_S\{|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|\le\beta_{CV}^{(n)}\}\ge1-\delta_{CV}^{(n)}.$$

The ontology learning map A is called distribution-independent, PH stable if for any positive integer n there exist a $\beta_{PH}^{(n)}$ such that for any $k\in\{1,\dots,n\}$ and any distribution μ, we infer $\lim_{n\to\infty}\beta_{PH}^{(n)}=0$ and

$$\mathbb{E}_S[|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|]\le\beta_{PH}^{(n)}.$$

The ontology learning map A is called distribution-independent, $Elto_{err}$ stable if for any positive integer n there exist a $\beta_{EL}^{(n)}$ and a $\delta_{EL}^{(n)}$ such that for any $k\in\{1,\dots,n\}$ and any distribution μ, we deduce $\lim_{n\to\infty}\beta_{EL}^{(n)}=0$, $\lim_{n\to\infty}\delta_{EL}^{(n)}=0$ and

$$\mathbb{P}_S\{|R(f_S)-\frac{1}{n}\sum_{k=1}^{n}L(f_{S^{i,j}},z_k)|\le\beta_{EL}^{(n)}\}\ge1-\delta_{EL}^{(n)}.$$

The ontology learning map A is called distribution-independent, leave-two-out hypothesis stable if for any positive integer n there exists a β_H^n such that for any distribution μ, we verify $\lim_{n\to\infty}\beta_H^n=0$ and

$$\mathbb{E}_{S,z}[|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|]\le\beta_H^n.$$

Furthermore, an ontology learning map A is called *LTO* stable if it shows both $CV_{\rm lto}$ and $Elto_{\rm err}$ stability.

The ontology learning map A is called distribution-independent, pseudo-pointwise hypothesis stable if for any positive integer n, $k\in\{1,\dots,n\}$ and any distribution μ, there exists a $\beta_{pPH}^{(n)}$ such that $\lim_{n\to\infty}\beta_{pPH}^{(n)}=0$ and

$$\mathbb{E}_S[L(f_{f_{S^{i,j}}},z_k)-L(f_{f_S},z_k)]\le\beta_{pPH}^{(n)}.$$

Pseudo-stability is also sufficient and necessary for universal consistency of ontology learning algorithm (judged by the conclusion in next section), but it is weaker than PH stability.

3. Main results and proofs

The results stated below manifest the equivalent of $CV_{\rm lto}$ stability and PH stability in ontology setting.

Theorem 1. *$CV_{\rm lto}$ stability with $\beta_{\rm lto}$ and $\delta_{\rm lto}$ leads to PH stability with $\beta_{PH}=\beta_{\rm lto}+M\delta_{\rm lto}$ and PH stability with β_{PH} reveals $CV_{\rm lto}$ stability with $(\alpha,\frac{\beta_{PH}}{\alpha})$ where $\alpha<\beta_{PH}$.*

Proof. In terms of the bound on the ontology loss function and the definition of $CV_{\rm lto}$ stability, we get for any $k\in\{1,\dots,n\}$,

$$\mathbb{E}_S[|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|]\le\beta_{\rm lto}+M\delta_{\rm lto}.$$

So, the first thesis is hold.

According to the definition of PH stability, we also obtain

$$\mathbb{E}_S[|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|]\le\beta_{PH}.$$

In view of Markov's inequality and the fact that $|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|\ge0$, we infer

$$\mathbb{P}[|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|>\alpha]$$
$$\le\frac{\mathbb{E}_S[|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|]}{\alpha}\le\frac{\beta_{PH}}{\alpha}.$$

Thus, we complete the proof. □

The next conclusion show that both $Elto_{\rm err}$ and $CV_{\rm lto}$ stability together are enough for generalization of symmetric ontology learning algorithms.

Theorem 2. *The generalization ontology learning error can be expressed as*

$$\mathbb{E}_S(R(f_S)-\widehat{R}_S(f_S))^2\le2\mathbb{E}_S(R(f_S)-\frac{\sum_{k=1}^{n}L(f_{S^{i,j},z_k})}{n})^2$$
$$+2M\mathbb{E}_s|L(f_{S^{i,j}},z_k)-L(f_S,z_k)|.$$

Proof. Obviously, the left hand part can be presented as

$$\mathbb{E}_S(R(f_S)-\widehat{R}_S(f_S))^2\ \le\ 2\mathbb{E}_S\left(R(f_S)-\frac{\sum_{k=1}^{n}L(f_{S^{i,j},z_k})}{n}\right)^2$$
$$+2\mathbb{E}_S\left(R(f_S)-\frac{\sum_{k=1}^{n}L(f_{S^{i,j},z_k})}{n}\right)^2.$$

The second term on the right hand can be further deduced as

$$\mathbb{E}_S\left(R(f_S) - \frac{\sum_{k=1}^n L(f_{S^{i,j}}, z_k)}{n}\right)^2$$
$$= \mathbb{E}_S\left(\frac{\sum_{k=1}^n L(f_S, z_k)}{n} - \frac{\sum_{k=1}^n L(f_{S^{i,j}}, z_k)}{n}\right)^2$$
$$= \mathbb{E}_S\sum_{k=1}^n |L(f_S, z_k) - L(f_{S^{i,j}}, z_k)|^2$$
$$\le M\mathbb{E}_S\sum_{k=1}^n |L(f_S, z_k) - L(f_{S^{i,j}}, z_k)|$$
$$\le M\mathbb{E}_S\sum_{k=1}^n |L(f_S, z_k) - L(f_{S^{i,j}}, z_k)|$$
$$= \frac{M}{n}\mathbb{E}_S\sum_{i=1}^n |L(f_S, z_i) - L(f_{S^{i,j}}, z_k)|$$
$$= M\mathbb{E}_S|L(f_S, z_k) - L(f_{S^{i,j}}, z_k)|.$$

□

From Theorem 2, we can directly get that LTO stability implies generalization in ontology setting.

From facts above, we can check the theorem below and the detailed proofs are skipped.

Theorem 3. *Suppose f_S, $f_{S^{i,j}} \in \mathcal{F}$ and ontology loss function is bounded. Then LTO stability is sufficient and necessary for consistency of ontology learning algorithm. Thus, the following statements are equivalent (i) the ontology map induced by almost ontology learning algorithm is LTO stable, (ii) almost ontology learning algorithm is universally consistent, (iii) $\mathcal{L}$ is uGC.*

The result above implies that $CV_{\rm lto}$ stability is sufficient and necessary for consistency of ontology learning algorithm on a function class $\mathcal{F}$ and ontology learning algorithm on a uGC class means $Elto_{\rm err}$ stability with $\lim_{n\to\beta}\beta = 0$ and

$$\mathbb{E}_S(R(f_S) - \frac{1}{n}\sum_{k=1}^n L(f_{S^{i,j}}, z_k))^2 \le \beta_n.$$

The lemmas presented next show the property of ε^E-minimizer.

Lemma 1. *There is a ε^E-minimizer that for any $k \in \{1, \ldots, n\}$*

$$L(f_S, z_k) - L(f_{S^{i,j}}, z_k) + 2(n-2)\varepsilon^E \ge 0.$$

Proof. According to definition of almost minimizer (1), we get

$$\frac{1}{n}\sum_{z_k\in S} L(f_{S^{i,j}}, z_k) - \frac{1}{n}\sum_{z_k\in S} L(f_S, z_k) \ge -\varepsilon_n^E,$$

$$\frac{1}{n}\sum_{z_k\in S^{i,j}} L(f_{S^{i,j}}, z_k) - \frac{1}{n}\sum_{z_k\in S^{i,j}} L(f_S, z_k) \le \frac{n-2}{n}\varepsilon_{n-2}^E,$$

The first inequality above can be re-formulated as

$$\left[\frac{1}{n}\sum_{z_k\in S^{i,j}} L(f_{S^{i,j}}, z_k) - \frac{1}{n}\sum_{z_k\in S^{i,j}} L(f_S, z_k)\right] + \frac{1}{n}L(f_{S^{i,j}}, z_k) - \frac{1}{n}L(f_S, z_k) \ge -\varepsilon_n^E.$$

Moreover, we get

$$L(f_{S^{i,j}}, z_k) - L(f_S, z_k) \ge -n\varepsilon_n^E - (n-2)\varepsilon_{n-2}^E.$$

By means of ε_n^N is a decreasing sequence, we infer

$$L(f_{S^{i,j}}, z_k) - L(f_S, z_k) \ge -2(n-2)\varepsilon_{n-2}^E.$$

Lemma 2. *In the almost ontology learning algorithm setting with $\varepsilon_n^E > 0$ selected with character $\lim_{n\to\infty} n\varepsilon_n^E = 0$, we get that for any $k \in \{1, \ldots, n\}$*

$$\mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|] \le \mathbb{E}_S R(f_{S^{i,j}}) - \mathbb{E}_S\widehat{R}_S(f_S) + 4(n-2)\varepsilon_{n-2}^E.$$

Proof. We observe that

$$\mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|]$$
$$= \mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k) + 2(n-2)\varepsilon_{n-2}^E - 2(n-2)\varepsilon_{n-2}^E|]$$
$$\le \mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k) + 2(n-2)\varepsilon_{n-2}^E|] + 2(n-2)\varepsilon_{n-2}^E.$$

Using the result in Lemma 1, we have that for any $k \in \{1, \ldots, n\}$,

$$L(f_S, z_k) - L(f_{S^{i,j}}, z_k) + 2(n-2)\varepsilon_{n-2}^E \ge 0.$$

Hence

$$\mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k) + 2(n-2)\varepsilon_{n-2}^E|] = \mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|] + 2(n-2)\varepsilon_{n-2}^E.$$

On the other hand, using the linearity of expectations, we have

$$\mathbb{E}_S[L(f_{S^{i,j}}, z_k) - L(f_S, z_k)] = \mathbb{E}_S R(f_{S^{i,j}}) - \mathbb{E}_S\widehat{R}_S(f_S)$$

and thus

$$\mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|] \le \mathbb{E}_S R(f_{S^{i,j}}) - \mathbb{E}_S\widehat{R}_S(f_S) + 4(n-2)\varepsilon_n^E.$$ □

Now, we get the fact on equivalent in ontology learning setting.

Theorem 4. *Assume the exact minimization of the ontology learning algorithm and the existence of the minima of the true risk $R(f^*)$ where $f* \in \arg\min_{f\in\mathcal{F}} R(f)$, then universal consistency is equivalent to (β, δ) $CV_{\rm lto}$ stability in ontology learning setting.*

Proof. According to the assumption, we get

$$L(f_{S^{i,j}}, z_k) - L(f_S, z_k) \ge 0.$$

Therefore, the following equivalences are established:

$$(\beta, \delta)CV_{\rm lto} \text{ stability} \Leftrightarrow \lim_{n\to\infty} \mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|] = 0$$
$$\Leftrightarrow \lim_{n\to\infty} \mathbb{E}_S[L(f_{S^{i,j}}, z_k) - L(f_S, z_k)] = 0$$

$$\Leftrightarrow \lim_{n\to\infty} \mathbb{E}_S R(f_{S^{i,j}}) - \mathbb{E}_S \widehat{R}_S(f_S) = 0$$
$$\Leftrightarrow \lim_{n\to\infty} \mathbb{E}_S R(f_{S^{i,j}}) = \lim_{n\to\infty} \mathbb{E}_S \widehat{R}_S(f_S).$$

In terms of $\widehat{R}_S(f_S) \le \widehat{R}_S(f^*)$ and $R(f^*) \le R(f_{S^{i,j}})$, we infer

$$R(f^*) \le \lim_{n\to\infty} \mathbb{E}_S R(f_{S^{i,j}}) = \lim_{n\to\infty} \mathbb{E}_S \widehat{R}_S(f_S)$$
$$\le \lim_{n\to\infty} \mathbb{E}_S \widehat{R}_S(f^*) = R(f^*),$$

and

$$\lim_{n\to\infty} \mathbb{E}_S R(f_{S^{i,j}}) = \lim_{n\to\infty} \mathbb{E}_S \widehat{R}_S(f_S) = \lim_{n\to\infty} \mathbb{E}_S \widehat{R}_S(f^*) = R(f^*),$$

which reveals that in probability $\lim_{n\to\infty} R(f_{S^{i,j}}) = R(f^*)$.

At last, the result is followed by the fact that the convergence of $R(f_S)$ to $R(f^*)$ is equivalent to the convergence of $R(f_{S^{i,j}})$ to $R(f^*)$ in probability. □

Theorem 5. *Assume that the ontology learning algorithm related on a class $\mathcal{F}$ is distribution-independent CV_{lto} stable, and the ontology loss is bounded by M. Then, the ontology learning algorithm on $\mathcal{F}$ is universally consistent.*

Proof. Let $S = (z_1, \ldots, z_n)$ and $S_{n+2} = \{z_1, \ldots, z_{n+2}\}$ are fixed ontology sample sets. According to the assumption of CV_{lto} stability for ontology learning algorithm, for $\beta_{PH}^{(n+2)} = \beta_{CV_{\text{lto}}}^{(n+2)} + M\delta(n+2)_{CV_{\text{lto}}}^{(n+2)}$ and any distribution μ, we have

$$\mathbb{E}_{S_{n+2}}[L(f_S, z_{n+2}) - L(f_{S_{n+2}}, z_{n+2})]$$
$$\le \mathbb{E}_{S_{n+2}}[|L(f_S, z_{n+2}) - L(f_{S_{n+2}}, z_{n+2})|] \le \beta_{PH}^{(n+2)}. \quad (4)$$

Furthermore, for any distribution μ, we infer

$$\mathbb{E}_S R(f_S) - \mathbb{E}_{S_{n+2}} \widehat{R}_{S_{n+2}}(f_{S_{n+2}})$$
$$= \mathbb{E}_{S_{n+2}}[L(f_S, z_{n+2}) - L(f_{S_{n+2}}, z_{n+2})]. \quad (5)$$

In view of (4) and (5), for any distribution μ, we infer

$$\mathbb{E}_S R(f_S) \le \mathbb{E}_{S_{n+2}} \widehat{R}_{S_{n+2}}[f_{S_{n+2}}] + \beta_{PH}^{(n+2)}. \quad (6)$$

Next, we show that

$$\lim_{n\to\infty} \sup_{\mu}(\mathbb{E}_S R(f_S) - \inf_{f\in\mathcal{F}} R(f)) = 0.$$

Let $\eta_\mu = \inf_{f\in\mathcal{F}} R(f)$ for distribution μ. Since $R(f) \ge \eta_\mu$ and $\mathbb{E}_S R(f_S) \ge \eta_\mu$ for any $f \in \mathcal{F}$, using (6), we obtain for any distribution μ,

$$\eta_\mu \le \mathbb{E}_S R(f_S) \le \mathbb{E}_S R(f_S) \le \mathbb{E}_{S_{n+2}} R_{S_{n+2}}(f_{\varepsilon_c,\mu}) + \beta_{PH}^{(n+2)}. \quad (7)$$

There exists $f_{\varepsilon_c,\mu} \in \mathcal{F}$ with $R(f_{\varepsilon_c,\mu}) < \eta_\mu + \varepsilon_c$ for any $\varepsilon_c > 0$. Using the property of almost ontology learning algorithm, we deduce

$$R_{S_{n+2}}(f_{S_{n+2}}) \le R_{S_{n+2}}(f_{\varepsilon_c,\mu}) + \varepsilon_{n+2}^E.$$

The following inequality (for any distribution μ) is obtained by taking expectations with respect to S_{n+2} and substituting in (7),

$$\eta_\mu \le \mathbb{E}_S R(f_S) \le \mathbb{E}_S R(f_S) \le \mathbb{E}_{S_{n+2}} \widehat{R}_{S_{n+2}}(f_{\varepsilon_c,\mu}) + \varepsilon_{n+2}^E + \beta_{PH}^{(n+2)}.$$

Fixed ontology function $f_{\varepsilon_c,\mu}$, in terms of $\lim_{n\to\infty} \varepsilon_{n+2}^E = 0$ and $\lim_{n\to\infty} \beta_{PH}^{(n+2)} = 0$, we get (for any distribution μ)

$$\mathbb{E}_{S_{n+2}} \widehat{R}_{S_{n+2}}[f_{\varepsilon_c,\mu}] = \frac{1}{n+2}\sum_{i=1}^{n+2} \mathbb{E}_{S_{n+2}} L(f_{\varepsilon_c,\mu}, z_i)$$
$$= R(f_{\varepsilon_c,\mu}) \le \eta_\mu + \varepsilon_c.$$

Assume n is a large number, for any distribution μ and any fixed $\varepsilon_c > 0$, we get

$$\eta_\mu \le \mathbb{E}_S R(f_S) \le \eta_\mu + \varepsilon_c.$$

Therefore, following from $\lim_{n\to\infty}\sup_\mu (R[f_S] - \eta_\mu) = 0$, for any $\varepsilon_c > 0$ we have

$$0 \le \lim_{n\to\infty} \inf \sup_\mu (R(f_S) - \eta_\mu) \le \lim_{n\to\infty} \sup \sup_\mu (R(f_S) - \eta_\mu) \le \varepsilon_c.$$

Let $V_S = R[f_S] - \eta_\mu$ be a random variable. Obviously, $V_S \ge 0$ and $\lim_{n\to\infty}\sup_\mu \mathbb{E}_S V_S = 0$. Following from Markov's inequality to V_S, we deduce that for any positive α,

$$\lim_{n\to\infty} \sup_\mu \mathbb{P}[R(f_S) > \eta_\mu + \alpha] \lim_{n\to\infty} \sup \mathbb{P}[V_S > \alpha]$$
$$\le \lim_{n\to\infty} \sup \frac{\mathbb{E}_S[V_S]}{\alpha} = 0.$$

This implies that given CV_{lto} stability, the distribution independent convergence of $R(f_S)$ to η_μ consistency in ontology learning setting. □

Theorem 6. *If the otology loss function is bounded, then consistency of ontology learning algorithm reveals CV_{lto} stability of ontology learning algorithm.*

Proof. Theorem 1 tells us that PH stability and CV_{lto} stability are equivalent if the ontology loss is bounded. For PH stability, we have to present that

$$\lim_{n\to\infty} \sup_\mu \mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|] = 0.$$

By means of Lemma 2, for any distribution μ, we get

$$\mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|]$$
$$\le \mathbb{E}_S R(f_{S^{i,j}}) - \mathbb{E}_S \widehat{R}_S(f_S) + 4(n-2)\varepsilon_{n-2}^E. \quad (8)$$

Fixed (universal) consistency, since a necessary and sufficient condition for universal consistency of ontology learning algorithm is that $\mathcal{L}$ is uGC, we obtain $\mathcal{L}$ is a uGC class and thus $R(f_{S^{i,j}})$ is close to $\widehat{R}_S(f_{S^{i,j}})$. Therefore, $R(f_{S^{i,j}}) - R_S(f_S)$ is small.

It is easy to get

$$\mathbb{E}_S[R(f_{S^{i,j}}) - \widehat{R}_S(f_S)]$$
$$= \mathbb{E}_S[R(f_{S^{i,j}}) - \widehat{R}_S(f_{S^{i,j}})] + \mathbb{E}_S[\widehat{R}_S(f_{S^{i,j}}) - \widehat{R}_S(f_S)]. \quad (9)$$

Using uGC property of $\mathcal{L}$ again, with probability at least $1 - \delta_n(\varepsilon_n)$, we obtain

$$R(f_{S^{i,j}}) - \widehat{R}_S(f_S) \le \varepsilon_n$$

and for any distribution μ

$$\mathbb{E}_S[\widehat{R}_S(f_{S^{i,j}}) - \widehat{R}_S(f_S)]$$
$$\le \mathbb{E}_S[|\widehat{R}_S(f_{S^{i,j}}) - \widehat{R}_S(f_S)|] \le \varepsilon_n + M\delta_n(\varepsilon_n). \quad (10)$$

In terms of

$$\widehat{R}_S(f_{S^{i,j}}) = \frac{(n-2)\widehat{R}_S(f_{S^{i,j}}) + L(f_{S^{i,j}}, z_k)}{n}$$
$$\le \frac{(n-2)(\widehat{R}_S(f_{S^{i}}) + \varepsilon_{n-2}^E) + L(f_{S^{i,j}}, z_k)}{n}$$

$$= \frac{(n-2)(\widehat{R}_S[f_{S^{i,j}}] + \varepsilon^E_{n-2}) + L(f_S, z_k) - L(f_S, z_k) + L(f_{S^{i,j}}, z_k)}{n} + \frac{n-2}{n}\varepsilon^E_{n-2} \le \widehat{R}_S(f_S) + \frac{M}{n} + \varepsilon^E_{n-2},$$

for any distribution μ, we infer

$$\mathbb{E}_S[\widehat{R}_S(f_{S^{i,j}}) - \widehat{R}_S(f_S)] \le \frac{M}{n} + \varepsilon^E_{n-2}. \tag{11}$$

Combining (9), (10) and (11), for any distribution μ, we obtain

$$\mathbb{E}_S[R(f_{S^i}) - \widehat{R}_S(f_S)] \le \varepsilon_n + M\delta_n(\varepsilon_n) + \frac{M}{n} + \varepsilon^E_{n-2}.$$

Using (8), for any distribution μ, we deduce

$$\mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|] \le \varepsilon_n + M\delta_n(\varepsilon_n) + \frac{M}{n} + \varepsilon^E_{n-2} + 4(n-2)\varepsilon^E_{n-2}.$$

Note that ε^E_n can be selected to be a decreasing sequence with $\lim_{n\to\infty}(4n-3)\varepsilon^E_n = 0$.

Further, if $\mathcal{L}$ is uGC, there exists a sequence $\varepsilon_n > 0$ such that $\lim_{n\to\infty}\varepsilon_n = 0$ and $\lim_{n\to\infty}\delta_n(\varepsilon_n) = 0$. Hence it is possible to select a sequence ε_n satisfies $\varepsilon_n \to 0$ and $\delta_n(\varepsilon_n) \to 0$. By combining these facts together, we get

$$\lim_{n\to\infty}\sup_{\mu}\mathbb{E}_S[|L(f_{S^{i,j}}, z_k) - L(f_S, z_k)|] = 0.$$

The desired result is proved since the universal consistency reveals PH hypothesis stability. □

The produce of proving implies that CV_{lto} stability means the leave-two-out error converges to the training error in probability.

Our last theorem shows that consistency of ontology learning algorithm implies $Elto_{\text{err}}$ stability.

Theorem 7. *Ontology learning algorithm on a uGC class reveals*

$$\mathbb{E}_S\left(R(f_S) - \frac{1}{n}\sum_{k=1}^{n} L(f_{S^{i,j}}, z_k)\right)^2 \le \beta_n,$$

where $\lim_{n\to\infty}\beta_n = 0$.

Proof. In terms of the triangle inequality, we have

$$\mathbb{E}_S\left(R(f_S) - \frac{1}{n}\sum_{k=1}^{n} L(f_{S^{i,j}}, z_k)\right)^2 \le 2\mathbb{E}_S(R(f_S) - \widehat{R}_S(f_S))^2 + 2\mathbb{E}_S\left(\widehat{R}_S(f_S) - \frac{1}{n}\sum_{k=1}^{n} L(f_{S^{i,j}}, z_k)\right)^2.$$

Since with probability $1-\delta_1$ we can check $|\widehat{R}_S(f_S) - R(f_S)| \le \beta_1$, thus

$$\mathbb{E}_S(\widehat{R}_S(f_S) - R(f_S))^2 \le M\beta_1 + M^2\delta_1.$$

On the other hand,

$$\mathbb{E}_S\left(\widehat{R}_S(f_S) - \frac{1}{n}\sum_{k=1}^{n} L(f_{S^{i,j}}, z_k)\right)^2 \le M\mathbb{E}_S|L(f_S, z_k) - L(f_{S^{i,j}}, z_k)|.$$

Since ontology learning algorithm is on a uGC class (β_2, δ_2) CV_{lto} stability establishes, we have

$$M\mathbb{E}_S|L(f_S, z_k) - L(f_{S^{i,j}}, z_k)| \le M\beta_2 + M^2\delta_2.$$

Therefore, we deduce

$$\mathbb{E}_S\left(\widehat{R}_S(f_S) - \frac{1}{n}\sum_{k=1}^{n} L(f_{S^{i,j}}, z_k)\right)^2 \le M\beta_2 + M^2\delta_2.$$

which lead to

$$\mathbb{E}_S\left(R(f_S) - \frac{1}{n}\sum_{k=1}^{n} L(f_{S^{i,j}}, z_k)\right)^2 \le 2M\beta_1 + 2M^2\delta_1 + 2M\beta_2 + 2M^2\delta_2.$$

□

4. Conclusion

Ontology, as a data structural storage, representation and computation model, has been employed in various subjects and been proved to have high efficiency. The problem of ontology learning algorithms is finding the similarity measure between concepts (vertices). Various learning techniques have been applied for ontology engineering in recent years. One popular learning trick is mapping each vertex v to a real number $f(v)$ by ontology function f, then the similarity between v_i and v_j is judged by $|f(v_i) - f(v_j)|$.

In this paper, we study the stability of ontology learning algorithm. The relationship between several stabilities are determined, and we present that leave-two-out stability is a necessary and sufficient condition for ontology learning algorithm.

References

[1] Przydzial BBJM, Koleti A. Gpcr ontology: development and application of a g protein-coupled receptor pharmacology knowledge framework. Bioinformatics 2013;29(24):3211–19. doi:10.1093/bioinformatics/btt565.
[2] Koehler SDS, C M. The human phenotype ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res 2014;42(D1):966–74. doi:10.1093/nar/gkt1026.
[3] Ivanovic M, Budimac Z. An overview of ontologies and data resources in medical domains. Expert Syst Appl 2014;41(11):5158–15166. doi:10.1016/j.eswa.2014.02.045.
[4] Hristoskova VSA, Zacharioudakis G. Ontology-driven monitoring of patient's vital signs enabling personalized medical detection and alert. Sensors 2014;14(1):1598–628. doi:10.3390/s140101598.
[5] Kabir JHMA, Yu J. User-centric social context information management: an ontology-based approach and platform. Pers Ubiquitous Comput 2014;18(5):1061–83. doi:10.1007/s00779-013-0720-9.
[6] Gao W, Xu T. Stability analysis of learning algorithms for ontology similarity computation. Abstr Appl Anal 2013;2013:9 pages. doi:10.1155/2013/174802.
[7] Gao W, Zhu L. Gradient learning algorithms for ontology computing. Comput Intell Neurosci 2014;2014:12 pages. doi:10.1155/2014/438291.