Journal of Applied Library and Information Science (JALIS)
Vol. 1, No. 1, 2012

www.worldsciencepublisher.org

Statistical Analysis for Ontology Algorithm

[1]Wei Gao, [1]Tianwei Xu, [2]Yun Gao

[1]School of Information and technology, Yunnan Normal University, Kunming, Yunnan 650500, China
[2]Editorial Department of Yunnan Normal University, Yunnan, Kunming 650092, China.

Email: gaowei@ynnu.edu.cn

Abstract –Ontology is an important topic in computer science. It has many applications in various fields. Ontology similarity computation plays a critical role in practical implementations. In this paper, we use an idea of applying scoring functions for ontology function from learning theory in the implementation of ontology similarity computation. Some statistical characters are given.

Keywords –Ontology, Computer network, Information retrieval, Search extension

1. Introduction

In information retrieval, ontology has been used to compute semantic similarity (see [1]) and search extensions for concepts. Every vertex on an ontology graph represents a concept; a user searches for a concept *A*, will return similarities concepts of *A* as search extensions. Let *G* be a graph corresponding to ontology *O*, the goal of ontology similarity measure is to approach a similarity function which maps each pair of vertices to a real number. Choose the parameter $M\in\mathbb{R}^+$, the concepts *A* and *B* have high similarity if Sim(*A*,*B*)>*M*. Choose the parameter $M\in\mathbb{R}^+$, let *A*,*B* be two concepts on ontology and Sim(*A*,*B*)>*M*, then return *B* as retrieval expand when search concept *A*. Therefore, the quality of similarity functions plays an important role in such applications. Moreover, ontology is also used in image retrieval (see [2-5]) in networks. Some effective methods for ontology similarity measure can be found in [6-11].

The common methods to design similarity functions for ontology applications are from the structure feature of ontology graph.

Example. Let $0 < \alpha, \beta, \gamma, \chi < 1$ be real numbers and $\alpha + \beta + \gamma + \chi = 1$. The similarity function can be represented as the weighting sum of their name similarity, structure similarity, instance similarly and attribute similarity:

$$Sim(A,B) = \alpha\, Sim_{\text{name}}(A,B) + \beta\, Sim_{\text{structe}}(A,B) + \gamma\, Sim_{\text{instance}}(A,B) + \chi\, Sim_{\text{attribute}}(A,B),$$

where each part of partial similarity is measured by name set, structure feature of ontology graph, instance set and attribute set of of two vertices, respectively. Thus, each partial similarity function may have complex formula and expression. Such methods have high computation complex and there are lots of parameters need to be chosen.

In recent years, the ontology problem has gained attention in machine learning. In such ontology algorithm, one learns a real-valued function that assigns scores to instances. What is important is the relative list of vertices induced by those scores. This problem is distinct from both classification and regression, and it is natural to ask what kinds of generalization properties hold for algorithms for this problem. Although there have been several recent advances in developing algorithms for various settings of the ontology problem, the study of generalization properties of ontology algorithms has been largely limited to the special setting.

Let *V* be the set of vertices of an ontology graph. For any two distinct vertices v_1 and v_2 it holds either $v_1 \prec v_2$ or $v_1 \succ v_2$, but does not known which is true. Assuming that real numbers y_1 and y_2 are assigned to the vertices v_1 and v_2 in ontology graph such that $v_1 \prec v_2$ is equivalent to $y_1 \le y_2$. Let *d*-dimensional vectors v_1 and v_2 (use vertex instead of its vector, and this vector contain all information for its vertices) describe observed or measured features of the vertices and let the observation space *V* be a Borel subset of $\mathbb{R}^d$. The aim of ontology algorithm on ontology graph is to obtain a ontology function $f: V\times V \to \mathbb{R}$, i.e., a ontology rule, in the following way:

if $f(v_1, v_2) \le 0$, then we predict that $y_1 \le y_2$.

It need introduce a probabilistic setting in order to measure the quality of a given ontology function *f*.. We assume that two vertices are randomly selected from the *V*. It is described by a pair of identically distributed and independent (with respect to the measure *P*) random vectors $Z_1 = (v_1, y_1)$ and $Z_2 = (v_2, y_2)$ taking values in $V\times\mathbb{R}$. Here, variables y_1 and y_2 define the ordering as above which are unknown.

A ontology loss function is a convex and nonnegative function *l* that assigns, for $f: V\times V\to\mathbb{R}$ and $v, v'\in V$, a non-negative real number $l(f; v, v_0)$ interpreted as the loss of *f* in its relative ontology of *v* and *v*'. Let

$$R_l(f) = \mathrm{E}(l(\mathrm{sign}(y_1 - y_2) f(v_1, v_2)))$$

be convex risk of a ontology rule f. That is, the expected ontology error on the ontology graph for a ontology function $f: V \times V \to \mathbb{R}$ associated with the ontology loss function l.

Denote the convex empirical risk as

$$\hat{R}_l(f) = \frac{1}{n(n-1)} \sum_{i \neq j} l_f(z_i, z_j)$$

where $l_f(z_i, z_j) = l(\mathrm{sign}(y_1 - y_2) f(v_1, v_2))$. Note that, for a fixed ontology ontology function f, $\hat{R}_l(f)$ is a U-statistic of the order two. Thus, features of U-process $\{\hat{R}_l(f) : f \in F\}$ are the basis for statistical characters of the ontology function $f_n = \arg\min_{f \in F} \hat{R}_l(f)$ as an estimator for unknown ontology function $f^* = \arg\min_{f \in F} R_l(f)$, where F is a family of ontology function.

We are interested in the excess risk of an estimator f_n for ontology applications. Generalization bounds are hot recent years in the learning theory, such research can refer [12-21].

In this paper, we give some statistical analysis for ontology learning algorithm.

2. Setting and Basics

Assume that ontology function is symmetric which implies $f(v_1, v_2) = - f(v_2, v_1)$ for every $f \in F$. And then, ontology function class F is uniformly bounded, i.e., there exists some constant $A_1 > 0$ such that for every $f \in F$ and each pair of vertices v_1, v_2 $\in V$, we get $|f(v_1, v_2)| \le A_1$. We will not reiterate these conditions again

Let μ be a probability measure on $V \times V$ and ρ_μ be a L^2-pseudo-metric on ontology function family F with

$$\rho_\mu(f_1, f_2) = \frac{1}{A_1} \sqrt{\int_{V \times V} [f_1(v_1, v_2) - f_2(v_1, v_2)]^2 d\mu(v_1, v_2)}.$$

The covering number $N(t, F, \rho_\mu)$ for ontology function family F with a radius $t > 0$ and a pseudo-metric ρ_μ is the minimal number of balls (with respect to ρ_μ) to cover F with centers in F and radii t needed. Therefore, $N(t, F, \rho_\mu)$ can be regard as the minimal number m with satisfies

$$\exists_{\bar{F} \subset F, |\bar{F}| = m} \ \forall_{f \in F} \ \exists_{\bar{f} \in \bar{F}} \ \rho_\mu(f, \bar{f}) \le t.$$

Think about the marginal distribution P^V for the vector V and following empirical measures: $P_n^V = \frac{1}{n} \sum_{i=1}^{n} \delta_{V_i}$ and $\nu_n = \frac{1}{n(n-1)} \sum_{i \neq j} \delta_{(V_i, V_j)}$ with counting measure $\delta_{(\cdot)}$. The ontology function family F that we discuss satisfies following assumption:

Assumption There exist constants D_i, $K_i > 0$, $i = 1,2$ such that for each measures of the form: $\mu_1 = P^V \otimes P_n^V$, $\mu_2 = \nu_n$ and every $t \in (0,1]$ we get

$$e^{N(t, F, \rho_\mu)} \le D_i t^{-K_i} \quad i = 1.2.$$

The tool used in this paper is Hoeffding's decomposition [22] for a U-statistic $\hat{R}_l(f) - \hat{R}_l(f^*)$ which allows to infer the equality

$$R_l(f) - R_l(f^*) - [\hat{R}_l(f) - \hat{R}_l(f^*)]$$
$$= 2P_n[R_l(f) - R_l(f^*) - P_{l_f} + P_{l_{f^*}}] - U_n(h_f - h_{f^*})$$

here

$$P_{l_f}(z_1) = E[l_f(Z_1, Z_2) | Z_1 = z_1],$$

$$P_n(g) = \frac{1}{n} \sum_{i=1}^{n} g(Z_i),$$

$$U_n(h_f - h_{f^*}) = \frac{1}{n(n-1)} \sum_{i \neq j} [h_f(Z_i, Z_j) - h_{f^*}(Z_i, Z_j)],$$

$$h_f(z_1, z_2) = l_f(z_1, z_2) - P_{l_f}(z_1) - P_{l_f}(z_2) + R_l(f).$$

Hence, a difference between a U-statistic $\hat{R}_l(f) - \hat{R}_l(f^*)$ and its expectation are break into the sum of iid random variables and a degenerate U-statistic $U_n(h_f - h_{f^*})$ by Hoeffding's decomposition. Such U-statistic degeneration implies the conditional expectation of its kernel is the zero-function, i.e., for each $z_1 \in V \times \mathbb{R}$, we have $E[h_f(Z_1, Z_2) - h_{f^*}(Z_1, Z_2) | Z_1 = z_1] = 0$.

Let Ω be a class of real ontology functions uniformly bounded by a constant $\Psi > 0$. And, $\varepsilon_1, \ldots, \varepsilon_n$ is Rademacher sequence of iid random variables. Variables ε_i's are independent of the sample $Z_1, \ldots, Z_n$ and take values in $\{1, -1\}$ with equal probability. With such Rademacher sequence, denote

$$\Delta_n(\Omega) = \sup_{\Omega \in \Psi} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i g(Z_i),$$

as an expression $E(\Delta_n(\Omega))$ the Rademacher average for class ψ. The expectation in such average is taken concern both samples $Z_1, \ldots, Z_n$ and $\varepsilon_1, \ldots, \varepsilon_n$.

Sub-root function is a non-decreasing and non-negative function ϕ: $[0, \infty) \to [0, \infty)$ satisfies that the function $r \to \frac{\phi(r)}{\sqrt{r}}$ is non-increasing for every $r > 0$. This sub-root function has a lot of good characters, for instance, continuous and has the unique positive fixed point r^* (it is the only positive solution for the equation $\phi(r) = r$). At last, we denote

$$\Omega^* = \{\alpha g : g \in \Omega, \alpha \in [0,1]\}$$

as a star-hull of $Pg = \mathrm{E}g(Z_1)$ and Ω.

Now, we introduce the following lemma which can be found in [23] and [24].

Lemma 1. [23, 24] Let the class Ω be such that for some constant $B>0$ and every $g \in \Omega$ we have $Pg^2 \le BPg$. Moreover, if there exists a sub-root function f with the fixed point r^*, which satisfies

$$\phi(r) \ge B\, E\Delta_n(g \in \Omega^* : Pg^2 \le r)$$

for each $r \ge r^*$, then for every $T > 1$ and $\alpha \in (0,1)$

$P(\ \forall_{g \in \Omega} \quad Pg \le \frac{T}{T-1} P_n(g) + \frac{6T}{B} r^* + [22 \ \Omega \ + 5BT]$

$\frac{\ln(1/\alpha)}{n}) \geq 1\text{-}\alpha$.

According to Lemma 1, we get our main result as follows.

Theorem 1. Let the family of ontology function F be convex and satisfy Assumption. If the modulus of convexity of a ontology loss function l have value in the interval $[-A_1, A_1]$ and with $\delta(t) \geq Ct^p$ for some constants $C>0$ and $p \leq 2$, then for each $\alpha \in (0,1)$ and $T>1$ with probability at least 1- α

$$\forall_{f\in F} \quad R_l(f) - R_l(f^*) \leq \frac{T}{T-1} P_n(Pl_f - Pl_{f^*}) + C_1 K_1 \frac{\ln n + \ln(1/\alpha)}{n}$$

where constant C_1 depend on T.

Proof. Consider the ontology functions family

$$Pl_f - Pl_{f^*} = \{ Pl_f - Pl_{f^*} : f \in F\}$$

Since ontology loss function l is convex, it is locally Lipschitz with constant L_f. Due to F is uniformly bounded, so $Pl_f - Pl_{f^*}$ is uniformly bounded by $2L_fA_1$ as well. Note that if the modulus of convexity of ontology loss l satisfies the assumption given in Theorem 1 and F is convex, then for some constant B and every function $f \in F$

$$E[Pl_f(Z_1) - Pl_{f^*}(Z_1)]^2 \leq B[\, R_l(f) - R_l(f^*)\,].$$

The precise value of the constant B can refer Lemma 5 of [25]. So, the relation required in Lemma 1 between expectations and second moments for functions from the considered class holds. Follow Lemma 1 for the class of ontology functions $\Omega = \{ \frac{Pl_f - Pl_{f^*}}{2L_l A_1} : f \in F\}$ and the sub-root function

$$\phi(r) = \frac{B}{2L_l A_1} E\Delta_n(g \in \Omega^* : Pg^2 \leq r)$$

we obtain the probabilistic inequality as follows

$$P(\forall_{f\in F} \quad R_l(f) - R_l(f^*) \leq \frac{T}{T-1} P_n(Pl_f - Pl_{f^*}) + C_1 r^* + C_2 \frac{\ln(1/\alpha)}{n}) \geq 1\text{-}\alpha .$$

Here C_1 and C_2 are constants.

Denote

$$\Omega_r^* = \{ g \in \Omega^* : Pg^2 \leq r \}$$

for certain $r > 0$ and

$$\xi = \sup_{g\in\Omega_r^*} \frac{1}{n} \sum_{i=1}^{n} g^2(Z_i) .$$

Applying Chaining Lemma for empirical processes in [25], we infer

$$E\Delta_n(\Omega_r^*) \leq \frac{C_1}{\sqrt{n}} E\int_0^{\frac{\sqrt{\xi}}{4}} \sqrt{\ln N(t, \Omega_r^*, \rho_{P_n})}\, dt , \quad (1)$$

where

$$\rho_{P_n}(g_1, g_2) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} [g_1(Z_i) - g_2(Z_i)]^2} .$$

Notice that $N(t, \Omega_r^*, \rho_{P_n}) \leq N(t, \Omega^*, \rho_{P_n}) \leq N(t/2, \Omega^*, \rho_{P_n})$. Moreover,

$$N(t, P_{l_f}, \rho_{P_n}) \leq N(t, P_{l_f}, \rho_{P\otimes P}) \leq N(\frac{t}{L_l}, F, \rho_{P^V \otimes P_n^V}) .$$

The first inequality above follows from Nolan and Pollard ([26], Lemma 20) and to verify the second one we use the fact that l is locally Lipschitz. Thus, Assumption above characters of covering numbers imply that for certain positive constants C and C_1

$$\ln N(t, \Omega_r^*, \rho_{P_n}) \leq C_1 K_1 \ln \frac{C}{t} .$$

Thus, the right hand of (1) can be bounded by $C_1\sqrt{\frac{K_1}{n}} E\int_0^{\frac{\sqrt{\xi}}{4}} \sqrt{\ln\frac{C}{t}}\, dt$. According to the Jensen's inequality, we infer

$$C_1\sqrt{\frac{K_1}{n}} E\int_0^{\frac{\sqrt{\xi}}{4}} \sqrt{\ln\frac{C}{t}}\, dt \leq C_1\sqrt{\frac{K_1}{n}} \sqrt{E\xi} \sqrt{\ln(\frac{C}{E\xi})}$$

Moreover, we get

$$E\xi \leq 8E\Delta_n(\Omega_r^*) + r$$

Combining what we discuss above, we obtain

$$E\Delta_n(\Omega_r^*) \leq C_1\sqrt{\frac{K_1}{n}} \sqrt{8E\Delta_n(\Omega_r^*) + r} \sqrt{\ln\frac{C}{t}} .$$

This show that for the fixed point r^* implies

$$r^* \leq C_1 K_1 \frac{\ln n}{n} . \qquad \square$$

Acknowledgements

First we thank the reviewers for their constructive comments in improving the quality of this paper. This work was supported in part by Key Laboratory of Educational Informatization for Nationalities, Ministry of Education, the National Natural Science Foundation of China (60903131) and Key Science and Technology Research Project of Education Ministry (210210). We also would like to thank the anonymous referees for providing us with constructive comments and suggestions.

References

[1] X. Su, J. Gulla, Semantic enrichment for ontology mapping, The 9th International Conference on Information Systems (NLDB), 2004, pp. 217-228.

[2] B. Hu, S. Dasmahapatra, P. Lewis, N. Shadbolt, Ontology-based medical image annotation with description logics, 15th IEEE International Conference on Tools with Artificial Intelligence, 2003, pp. 77-82.

[3] S. Liu , L. Chia, S. Chan, Ontology for naturescene image retrieval, In on the move to meaningful Internet systems, CoopIS, DOA, and ODBASE, 2004, pp. 1050-1061.

[4] J. Opitz, B. Parsia, U. Sattler, Using Ontologies for Medical Image Retrieval- An Experiment, Proceedings 5th Int. Workshop on OWL: Experiences and Directions, 2009, pp. 23-24.

[5] H. Wang, S. Liu, L. Chia, Does ontology help in image retrieval?- A comparison between keyword, text ontology and multi-modality

ontology approaches, MM06, October, Santa Barbara, California, USA. ACM, 2006, pp. 109-112.

[6] W. Gao, L. Liang, Ontology Similarity Measure by Optimizing NDCG Measure and Application in Physics Education, Future Communication, Computing, Control and Management, LNEE 142, 2011, pp. 415–421.

[7] Y. Gao, W. Gao, Ontology Similarity Measure and Ontology Mapping via Learning Optimization Similarity Function, International Journal of Machine Learning and Computing, 2012, vol. 2, no. 2, pp.107-112.

[8] X. Huang, T. Xu, W. Gao, Z. Jia, Ontology Similarity Measure and Ontology Mapping Via Fast Ontology Method, International Journal of Applied Physics and Mathematics, 2011, vol. 1, no. 1, pp.54-59.

[9] M. Lan, Y. Ren, J. Xu, W. Gao, Ontology similarity computation using k-partite ontology method, Journal of Computer Applications, 2012, vol.32, no.4, pp.1094-1096.

[10] W. Gao, M. Lan, Ontology mapping algorithm based on ontology learning method, Microelectronics & computer, 2011, vol. 28, no.9, pp.59-61.

[11] Y. Wang, W. Gao, Y. Zhang, and Y. Gao, “Ontology Similarity Computation Use Ontology Learning Method”, The 3rd International Conference on Computational Intelligence and Industrial Application, 2010, pp. 20-22.

[12] X. He, W Gao, and Z Jia, Generalization Bounds of Ontology via Query-Level Stability I, ICCIC 2011, Part VI, CCIS 236, 2011, pp.188–196.

[13] Z. Jia, W Gao, X He, Generalization Bounds for Ontology Algorithm via Query-Level Stabilities Analysis, ICCIC 2011, Part VI, CCIS 236, 2011, pp. 197–203.

[14] Y. Gao, W. Gao, Y. Zhang, Query-Level Stability of IRSVM for Replacement Case, Procedia Engineering, 2011, vol. 15, pp. 2150 – 2154.

[15] Y. Gao, W. Gao, Y. Zhang, Query-Level Stability of Ontology SVM for Replacement Case, Procedia Engineering, vol. 15, 2011, pp. 2145– 2149.

[16] W. Gao, Y. Zhang, Generalization Bounds for Certain Class of Ontology Algorithm, Advanced Materials Research, 2011, vol. 267, pp. 456-461.

[17] W. Gao, Y. Zhang, L. Liang, Stability Analysis for Local Transductive Regression Algorithms, Advanced Materials Research, 2011, vol. 267, pp. 438-443.

[18] W. Gao, Y. Zhang, L. Liang, Y. Xia, Stability Analysis for Ontology Algorithms, In Proceedings 2010 IEEE International Conference on Information Theory and Information Security, Beijing, China, Dec, 2010, pp.973-976.

[19] W. Gao, Y. Zhang, Y. Gao, L. Liang, Y. Xia, Strong and Weak Stability of Bipartite Ontology Algorithms, International Conference on Engineering and Information Management, April, 2011, Chengdu, China, IEEE press, pp.303-307.

[20] T. Xu, J. Gan, Y. Zhang, Y. Gao, Generalization Bounds for Ontology Algorithms via Almost Everywhere Stability, 2011 3rd IEEE international conference on computer research and development, March, 2011, Shanghai, China, IEEE press, Vol, 04, pp.500-503.

[21] T. Xu, T. Zhang, W. Gao, Generalization Bounds for Ordinal Regression Algorithms via Strong and Weak Stability, 2011 3rd International conference on Machine learning and computing, Singapore, Feb, 2011: 143-147.

[22] V. H. de la Pe˜na and E. Gin´e. Decoupling: From Dependence to Independence. Springer-Verlag, New York, 1999.

[23] P. Massart, Some applications of concentration inequalities to statistics. Probability theory, Ann. Fac. Sci. Toulouse Math., 2000, vol. 9, pp. 245–303.

[24] P. L. Bartlett, O. Bousquet, S. Mendelson, Local rademacher complexities, Ann. Statist., 2005, vol. 33, pp. 1497–1537.

[25] D. Pollard, Convergence of Stochastic Processes. Springer, New York, 1984.

[26] D. Nolan, D. Pollard, U-processes: rates of convergence. Ann. Statist., 1987, vol. 15, pp. 780–799.

[27] W. Rejchel, On Ontology and Generalization Bounds, Journal of Machine Learning Research, 2012, vol. 13, pp.1373-1392.

[28] W. Gao, C. Yan, A note on fractional (k,n',m)-critical deleted graph, Advances in Computational Mathematics and its Applications, 2012, Vol. 1, No. 1, pp. 53-55.

[29] W. Gao, Y. Wang, Note on some characters for fractional (g,f,n)-critical graphs, Advances in Intelligent Transportation Systems (AITS), 2012, Vol. 1, No. 3, pp. 75-77.

[30] W. Gao, Y. Wang, Z. Zhang, One degree condition for graph G to be a fractional ID-(k,m)-deleted graph, Journal of Science (JOS), 2012, Vol. 1, No. 2, pp. 3-5.

[31] X. Huang, T. Xu, W. Gao, S. Gong, Ontology Similarity Measure and Ontology Mapping Using Half Transductive Ontology, Journal of Engineering (JOE), 2012, Vol. 1, No. 2, pp. 20-24.

Vitae

Wei Gao, male, was born in the city of Shaoxing, Zhejiang Province, China on Feb.13, 1981. He got two bachelor degrees on computer science from Zhejiang industrial university in 2004 and mathematics education form College of Zhejiang education in 2006. Then, he enrolled in department of computer science and information technology, Yunnan normal university, and got Master degree there in 2009. In 2012, he got PHD degree in department of Mathematics, Soochow University, China.

Now, he acts as lecturer in the department of information, Yunnan Normal University. As a researcher in computer science and mathematics, his interests are covered two disciplines: Graph theory, Statistical learning theory, Information retrieval, and Artificial intelligence.

Tianwei Xu, male, was born in Maguan City, Yunnan Province, China in 1970. He got bachelor degree on physics from the department of physics, Yunnan normal university, 1992. Then, he got Master degree of computer science major from the Yunnan Normal University in June, 2006. Now he is PhD student in the institute of higher education, Huazhong University of science and technology.

Act as a lecturer in the department of physics, Yunnan normal university, from July 1992 to June 1997. Then, act as associate secretary in the department of computer science and information technology, Yunnan normal university, from July 1997 to June 2003, and associate minister in the department of organization, Yunnan normal university, from July 2003 to June 2007. He also visited Curtin University of Technology in Austrian from Dec 2005 to June 2006. From 2007, he worked as secretary and associate professor in the department of information, Yunnan normal university, and worked as professor from 2011. He is also a PHD student in the institute of higher education, Huazhong University of science and technology. He also acts as leader in the Key Laboratory of National Education Informatization, Ministry of Education, Yunnan Normal University, China.

Yun Gao, male, was born in Xingping, Yunnan Province ,China, on Dec.10,1964. Studied in the department of physics of Ynnan Normal University from 1980 to 1984 and got bachelor degree on physics. And then worked in the Journal of Yunnan Normal University as an editor and physics lecturer till now. During this years, always engaged in studying the computational physics and other related areas, such as condensed matter physics and computer science, have published more than 30 scientific papers in home and abroad.

Now, as the vice editor in chief of Yunnan Normal University and a researcher, his interests are mainly in: computational physics and computing method.