SAS Journal of Medicine ISSN 2454-5112
SAS J. Med., Volume-1; Issue-3 (Sep-Oct, 2015); p-99-101
Available online at http://sassociety.com/sasjm/

Short Communication

Short Communication on Regression Based Ontology Algorithm

Yun Gao[1], Wei Gao[2]
[1]Department of Editorial, Yunnan Normal University, Kunming 650092, China
[2]School of Information, Yunnan Normal University, Kunming, Yunnan 650500, China

***Corresponding author**
Wei Gao
Email: gaowei@ynnu.edu.cn

Abstract: As a powerful tool, ontology even employed in various subjects, such as neuroscience, medical science, pharmacopedia, chemistry, education and other social science. In this short communication, we report an ontology algorithm using the special loss function.
Keywords: ontology, similarity measure, ontology mapping, loss function

INTRODUCTION

Ontology, a knowledge representation and conceptual shared model, has proved itself to be useful in image retrieval, knowledge management and information retrieval search extension. What's more, as an effective concept semantic model, ontology also finds its place in the other disciplines like social science, medical science, biology science, pharmacology science and geography science. Actually, the ontology model is a graph $G=(V,E)$, each vertex v in an ontology graph G stands for a concept and each edge $e=v_iv_j$ of an ontology graph G stands for a relationship between concepts v_i and v_j .

Concerning ontology similarity measure and ontology mapping, several effective learning tricks work well. Wanget al., [1] tended to learn a score function to map each vertex to a real number. Then, according to the difference of the real number which the two vertices correspond to, we can measure the similarity between them. Huang *et al.*, [2] worked out a fast ontology algorithm to calculate the ontology similarity within a short time. Gao and Liang [3] reported that the optimal ontology function can be determined by optimizing NDCG measure. And they also took the idea to physics education. Gao and Gao [4] deduced the ontology function through the regression approach. Moreover, based on half transductive learning, Huang *et al.*, [5] obtainedontology similarity function. Gao *et al.*, [6] raisednew ontology mapping algorithm by means of harmonicanalysis and diffusion regularization on hypergraph. Gao and Shi [7] proposed new ontology similarity computation technology. As a result, the new calculation model considering operational cost in the real implement. Few years ago, Gao and Xu [8] presented the ontology similarity measuring and ontology mapping algorithms on basis of minimum error entropy criterion. Several theoretical analysis of ontology algorithm can refer to Gao *et al.*, [9], Gao and Xu [10],Gao and Zhu [11] and Gao *et al.*, [12].

Gao and Gao [4] introduced the regression based ontology learning framework, and we continue employ this framework here. The different between [4] and our short communication is that we use the special loss function in the regression framework.

Algorithm Description

Let V be an instance space. For each vertex in ontology graph, a p dimension vector expresses information including its name, instance, attribute and structure, and semantic information of the concept which corresponds to the vertex and that is contained in name and attributes components of its vector. To promote the representation, we try confusing the notations and using v to denote both the ontology vertex and its corresponding vector. The ontology learning algorithm s are set to get an optimal ontology (score) function $f:V\rightarrow \mathbb{R}$, and the similarity between two vertices is judged by the difference between two corresponding real numbers.

Correntropy is a generalized similarity measure between two scalar random variables U_1 and U_2, which is defined by $\Theta_\sigma(U_1,U_2)=\mathrm{E}K_\sigma(U_1,U_2)$. Here K_σ is a Gaussian kernel given by $K_\sigma(u_1,u_2)=e^{-\frac{(u_1-u_2)^2}{\sigma^2}}$ with the scale parameter $\sigma>0$, (u_1,u_2) being a realization of (U_1,U_2). For given ontology data set $\mathbf{z}=\{(v_i,y_i)\}_{i=1}^m$, the maximum correntropy criterion for regression models the output ontology function via maximizing the empirical estimator of Θ_σ as follows:

$$f_z=\arg\max_{f\in H}\frac{1}{m}\Theta_\sigma(y_i,f(v_i)),$$

Where *H* is a ontology function space.

The correntropy induced regression loss $l_\sigma:\mathbb{R}\times\mathbb{R}\to[0,+\infty)$ is defined as

$$l_\sigma(y,t)=\sigma^2(1-e^{-\frac{(y-t)^2}{\sigma^2}}),\quad y\in Y,t\in\mathbb{R},$$

with $\sigma>0$ being a scale parameter. We use correntropy induced regression loss to the ontology framework described in [4], then the ontology algorithm becomes

$$f_z=\arg\max_{f\in H}\frac{1}{m}l_\sigma(y_i,f(v_i)), \qquad (1)$$

where the ontology space *H* is assumed to be a compact subset of *C*(*V*).

Simulation Studies

In this section, we designed two simulation experiments which are related to ontology similarity measure and ontology mapping, respectively.

First, we use "PO" ontology (constructed in http: //www.plantontology.org.) to check the efficiency of our ontology algorithm (1)in ontology similarity measuring. *P@N* standard [13] is used for this experiment. Taking *N*=3, 5 and 10, the results are *P*@3=37.51%, *P*@5=39.11% and *P*@10=60.32%. Thus, the ontology algorithm (1) has high effective on plant science data.

Second, we use physics ontology (constructed in [3]) to check the efficiency of our ontology algorithm (1) in ontology mapping. *P@N* standard is also used for this experiment. Taking *N*=1, 3 and 5, the results are *P*@1=35.48%, *P*@3=44.09% and *P*@5=66.45%. Thus, the ontology algorithm (1) has high effective on physics data.

DISCUSSION

If the ontology function is assumed to be linear, i.e., $f(v)=v^{\mathrm{T}}\theta$ for some linear vector $\theta\in\mathbb{R}^p$. Then, the ontology optimization problem can be stated as

$$\hat{\theta}=\arg\min_{\theta\in\mathbb{R}^p}\sum_{i=1}^m\phi\left(\frac{y_i-v_i^{\mathrm{T}}\theta}{\sigma}\right),$$

where $\sigma>0$ is the scale parameter and ϕ is a robust loss function that down weights large residual errors. In fact, by using the above robust loss function ϕ, we have the following robust nonparametric ERM-based ontology regression scheme

$$\hat{f}_z=\arg\min_{f\in H}\sum_{i=1}^m\phi\left(\frac{y_i-f(v_i)}{\sigma}\right).$$

Denoting $\phi_\sigma(t)=\phi(\frac{t}{\sigma})$, besides the l_σ ontology loss, several frequently employed robust ontology loss functions include:

- Huber's loss: $\phi_\sigma(t)=t^2I\{|t|\le\sigma\}+(2\sigma|t|-\sigma^2)I\{|t|>\sigma\}$;
- Cauchy loss: $\phi_\sigma(t)=\sigma^2\log(1+\frac{t^2}{\sigma^2})$;

- Tukey's biweight loss: $\phi_\sigma(t) = \frac{\sigma^2}{6}(1-(1-(\frac{t}{\theta})^2)^3)I\{|t| \le \sigma\} + \frac{\sigma^2}{6}I\{|t| > \sigma\}$.

REFERENCE

1. Wang Y, Gao W, Zhang Y, Gao Y; Ontology similarity computation use ranking learning Method, In: The 3rd International Conference on Computational Intelligence and Industrial Application, Wuhan, China, 2010; 20-22.
2. Huang X, Xu T, Gao W, Jia Z; Ontology similarity measure and ontology mapping via fast ranking method, International Journal of Applied Physics and Mathematics, 2011; 1(1): 54-59.
3. Gao W, Liang L; Ontology similarity measure by optimizing NDCG measure and application in physics education," Future Communication, Computing, Control and Management, 2011; 142: 415-421.
4. Gao Y, Gao W; Ontology similarity measure and ontology mappingvia learning optimization similarity function, International Journal of Machine Learning and Computing, 2012; 2(2):107-112.
5. Huang X, Xu T, Gao W, Gong S; Ontology similarity measure and ontology mapping using half transductive ranking," In: Proceedings of 2011 4th IEEE international conference on computer science and Information technology, Chengdu, China, 2011;571-574.
6. Gao W, Gao Y, Liang L; Diffusion and harmonic analysison hypergraph and application in ontology similarity measure and ontology mapping, Journal of Chemical and Pharmaceutical Research, 2013; 5(9): 592-598.
7. Gao W, Shi L; Ontology similarity measure algorithm with operational cost and application in biology science, Bio Technology:An Indian Journal, 2013; 8(11): 1572-1577.
8. Gao W, Xu TW; Ontology similarity measuring and ontology mapping algorithm based on MEE criterion, Energy Education Science and Technology Part A: Energy Science and Research, 2014: 32(5): 3793-3806.
9. Gao W, Gao Y, Zhang Y; Strong and weak stability of k-partite ranking algorithm, Information, 2012; 15(11): 4585-4590.
10. Gao W, Xu TW; Stability analysis of learning algorithms for ontology similarity computation," Abstract and Applied Analysis, 2013; 9.
11. Gao W, Zhu L; Gradient learning algorithms for ontology computing, Computational Intelligence and Neuroscience, 2014; 12.
12. Gao W, Yan L, and Liang L; Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithmin AUC criterion setting, International Journal of Computer. Applications in Technology, 2014; 50: 226-231.
13. Craswell N, Hawkin D; Overview of the TREC 2003 webtrack, In: Proceeding of the 12th Text Retrieval Conference, Gaithersburg, Maryland, 2003; 78-92.