The Open Cybernetics & Systemics Journal

BENTHAM OPEN

Ontology Sparse Vector Learning Based on Accelerated First-Order Method

Yun Gao and Wei Gao

Ontology Sparse Vector Learning Based on Accelerated First-Order Method

Yun Gao[1], Wei Gao[2*]

[1]*Department of Editorial, Yunnan Normal University, Kunming 650092, China*

[2] *School of Information and Technology, Yunnan Normal University, Kunming 650500, China*

* Corresponding author: gaowei@ynnu.edu.cn

Abstract: In this article, we present a sparse vector learning algorithm for ontology similarity measure and ontology mapping by virtue of accelerated first-order technology. The main procedure of our iterative algorithm is relying on proximity operator computation, Picard-Opial process and accelerated first-order tricks. The simulation experimental results show that the new proposed algorithm has high efficiency and accuracy in ontology similarity measure and ontology mapping in plant science and university application.

Keywords: Ontology, similarity measure, ontology mapping, sparse vector, accelerated first-order method

1. 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 (for instance, see Przydzial et al., [1], Koehler et al., [2], Ivanovic and Budimac [3], Hristoskova et al., [4], and Kabir et al., [5]).

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. Ontology similarity-based technologies were employed in a variety of applications several years ago. On the condition of technology for stable semantic measurement, a graph derivation representation which is based on trick for stable semantic measurement is proposed by Ma et al., [6]. Later,

Li et al., [7] contributed to pointing out an ontology representation method which can be used for online shopping customers knowledge in enterprise information. Santodomingo et al., [8] explained a creative ontology matching system. In his view, the complex correspondences are presented by processing expert knowledge from external domain ontologies and in terms of using novel matching tricks. Pizzuti et al., [9] gave a detailed description about the main features of the food ontology and some examples of application for traceability purposes. Lasierra et al., [10] proposed that ontologies are very useful in designing an architecture for looking after patients at home.

Concerning ontology similarity measure and ontology mapping, several effective learning tricks work well. Wang et al., [11] 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., [12] worked out a fast ontology algorithm to calculate the ontology similarity

within a short time. Gao and Liang [13] 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 [14] deduced the ontology function through the regression approach. Moreover, based on half transductive learning, Huang et al., [15] obtained ontology similarity function. Gao et al., [16] raised new ontology mapping algorithm by means of harmonic analysis and diffusion regularization on hypergraph. Gao and Shi [17] 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 [18] 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., [19], Gao and Xu [20], Gao and Zhu [21] and Gao et al., [22].

In this paper, we present the accelerated first-order based ontology algorithm for ontology similarity computation and ontology mapping. By means of the sparse vector, the ontology graph is mapped into a real line and vertices are mapped into real numbers. Then the similarity between vertices is measured by the difference between their corresponding real numbers. The rest of the paper is arranged as follows: we present the notations and setting in Section 2; the ontology optimization algorithm and iterative strategies are raised in Section 3; at last, the experiments on plant ontology and university ontology are designed to show the efficiency of the algorithm.

2. SETTING

Let V be a 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. Let $v = \{v_1, \cdots, v_p\}$ be a vector that corresponds to a vertex v. 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 algorithms are set to get an optimal ontology (score) function $f: V \rightarrow$, and the similarity between two vertices is judged by the difference between two corresponding real numbers. The core of this algorithm is dimensionality reduction, i.e., choosing one dimension vector to express p dimension vector. In specific, an ontology function f is a dimensionality reduction function f:

$$^p \rightarrow \quad .$$

In the practice implement, one sparse ontology function is expressed by

$$f_\beta(v) = \sum_{i=1}^{p} v_i \beta_i + \delta . \qquad (1)$$

Here $\hat{a} = (\beta_1, \cdots, \beta_p)$ is a sparse vector and δ is a noise term. The sparse vector $\hat{a}$ is to shrink irrelevant component to zero. To determine the ontology function f, we should learn the sparse vector $\hat{a}$ first.

In our paper, we consider the general versions for learning $\hat{a}$. Let $\{v_i, y_i\}_{i=1}^{n}$ be a sample set with n vertex, $V \in ^{n \times p}$ be the matrix of n samples such that each sample vertex lies in a p dimension space, and $y = (y_1, \cdots, y_n) \in ^n$ be the vector of outputs of the these n sample vertex. Hence, the regression function (1) can be expressed as the linear model:

$$y = V\hat{a} + \delta ,$$

(2)

where δ is the n dimension vector for noise

distributed as $N(0, \sigma^2 I_{n \times n})$.

The general regression obtains an estimate of the sparse vector by solving the following optimization problem:

$$\min_{\hat{a} \in {}^p} l(\hat{a}) + \lambda \|\hat{a}\|_1, \qquad (3)$$

where $l(\hat{a}) = \frac{1}{2} \|y - V\hat{a}\|_2^2$ is the loss term, $\|\hat{a}\|_1 = \sum_{i=1}^p |\beta_i|$ is the l_1-norm balance term that measures the sparseness of vector $\hat{a}$, and $\lambda > 0$ is the balance parameter which controls the sparsity level.

In what follows, let $|\hat{a}| = (|\beta_i|)_{i=1}^p \in {}^p$ be the vector. For each $q \geq 1$, the l_q-norm of sparse vector $\hat{a}$ is denoted by $\|\hat{a}\|_q = (\sum_{i=1}^p |\hat{a}_i|^q)^{\frac{1}{q}}$.

Let $\delta_C : {}^p \to$ be the indicator function of the set $C \subseteq {}^p$ such that $\delta_C(x) = 0$ if $x \in C$ and $\delta_C(x) = +\infty$ otherwise. Furthermore, let $\partial \varphi(x)$ be the sub-differential of φ at $x \in {}^p$ where $\varphi : {}^p \to$ is a convex function.

3. MAIN ONTOLOGY ALGORITHM AND SOLUTION TECHNOLOGIES
3.1 Optimization Framework

Let $\rho > 0$ be a parameter, Λ be a prescribed convex subset, and $\Gamma : {}^p \times {}^p_+ \to$ be a

function given by $\Gamma(\beta, \lambda) = \frac{1}{2} \sum_{i=1}^p (\frac{\beta_i^2}{\lambda_i} + \lambda_i)$.

Then, our ontology optimization model can be expressed as

$$\inf\{\frac{1}{2} \|V\beta - y\|_2^2 + \rho \Gamma(\beta, \lambda) : \beta \in {}^p, \lambda \in \Lambda\}. (4)$$

Here, λ is an auxiliary vector. Above framework can be further formulated by

$$\min\{\frac{1}{2} \|V\beta - y\|_2^2 + \rho \Omega(\beta) : \beta \in {}^p\} \qquad (5)$$

where the balance term takes the form

$$\Omega(\beta) = \inf\{\Gamma(\beta, \lambda) : \lambda \in \Lambda\}.$$

Since Γ is jointly convex, the framework (5) is still convex. Moreover, the balance function Ω is independent of the signs of the elements of ontology sparse vector $\hat{a}$.

Now, we introduce proximity operator for optimizing our ontology model. Let φ be a real-valued convex function on ^{d}. For each $t \in {}^d$, the proximity operator of φ is denoted by

$$\text{prox}_{\varphi}(t) = \arg\min\{\frac{1}{2} \|z - t\|_2^2 + \varphi(z) : z \in {}^d\}.$$

This operator is well-defined since the above minimum exists and unique.

3.2 Selection of Generic Convex Set Λ

Now, we focus on the generic convex set Λ. For each $\lambda \in {}^p_+$, the quadratic function $\Gamma(\cdot, \lambda)$ supplies an upper bound to the l_1-norm in view of $\Omega(\hat{a}) \geq \|\hat{a}\|_1$, and the equality is hold if and only if $|\hat{a}| \in \Lambda$ according to the arithmetic-geometric inequality.

For each solution $(\hat{a}, \hat{\lambda})$, the sparsity pattern

of $\hat{a}$ is included in the sparsity pattern of $\hat{\lambda}$, i.e., the indices associated with nonzero components of $\hat{a}$ are a subset of those of $\hat{\lambda}$. If $\hat{\lambda} = 0$, then

$\hat{a} = 0$ also holds in view of the ratio $\dfrac{\beta_i^2}{\lambda_i}$. Thus,

if the set Λ supports some sparse solutions of $\hat{\lambda}$, then the same sparsity pattern will be manifested on $\hat{a}$. Furthermore, the sparse vectors λ is supported by $\sum_i \lambda_i$. For instance, a restriction of the form $\lambda_1 \geq \lambda_2 \geq \ldots \cdot \geq \lambda_n$ supports consecutive zeros at the last elements of λ and non-zeros in other positions. As a result, it lead to zeros at the last elements of $\hat{a}$. In this point of view, in many situations like this, it is not hard to combine a convex constraint on λ, but it may not be available to do the same directly on $\hat{a}$.

In our paper, the sets Λ we considered have the expression as follows

$$\Lambda = \{\lambda \in \ _{++}^p : A\lambda = S\}$$

where A is a $k \times p$ matrix and S is a convex set.

Two main selections of interest are when S is the unit ball of a norm or a convex cone. We refer to the corresponding set Λ as norm constraint set or conic constraint, respectively. Two specific examples are considered to respect the sparsity patterns with each selection.

For conic constraint sets, we select $S = \ _+^k$, and thus $\Lambda = \{\lambda \in \ _+^p : A\lambda \geq 0\}$. For the norm constraint sets, we select S as the l_1-unit ball and

Λ is the edge map of an ontology graph G with edge set E, i.e., $\Lambda = \{\lambda \in \ _+^p : \sum_{(v_i,v_j) \in E} |\lambda_i - \lambda_j| \leq 1\}$.

Furthermore, if G is a one dimension grid ontology graph, then it can be expressed as

$$\Lambda = \{\lambda \in \ _+^p : \sum_{i=1}^{p-1} |\lambda_i - \lambda_j| \leq 1\}.$$

3.3 Calculating the Proximity Operator

In this subsection, we give the strategies to solve problem (4) using the accelerated first-order technologies. This tricks depend on the proximity operator computation restricted to $\ ^p \times \Lambda$.

By virtue of its definition, the proximal operator of Γ at $(\alpha, \mu) \in \ ^P \times \ ^P$ is the solution of the following ontology problem

$$\inf\{\frac{1}{2}\|(\beta,\lambda) - (\alpha,\mu)\|_2^2 +$$

$$\rho\Gamma(\beta,\lambda) : \beta \in \ ^p, \lambda \in \Lambda\}. \qquad (6)$$

For given λ, a calculating gets that the objective function in (6) reaches its minimum at

$$\beta_i(\lambda) = \frac{\alpha_i \lambda_i}{\lambda_i + \rho}.$$

In terms of this fact, we infer the modified ontology problem

$$\min\{\frac{1}{2}\|\lambda - \mu\|_2^2 + \frac{\rho}{2}\sum_{i=1}^p (\frac{\alpha_i^2}{\lambda_i + \rho} + \lambda_i) : \lambda \in \Lambda\}. (7)$$

We consider how to solve (7) in our general setting $\Lambda = \{\lambda : \lambda \in \ _+^p : A\lambda \in S\}$. Suppose that the projection on the set S can be smoothly determined. For this purpose, let $B = \begin{bmatrix} I \\ A \end{bmatrix}$ be

the $(p+k) \times p$ matrix and the function

$$\varphi(s,t) = \varphi_1(s) + \varphi_2(t), \text{ where } (s,t) \in$$

$^p \times {}^k$, $\varphi_2(t) = \delta_S(t)$ and

$$\varphi_1(s) = \frac{\rho}{2} \sum_{i=1}^{p} (\frac{\alpha_i^2}{s_i + \rho} + s_i + \delta_{++}(s_i)).$$

One fact we emphasize here that the solution of problem (7) is the same as the proximity function of the linearly composite map $\varphi \circ B$ at μ, which gets the solution of the problem

$$\min\{\frac{1}{2}\|\lambda - \mu\|^2 + \varphi(B\lambda) : \lambda \in {}^p\}.$$

3.4 Algorithm presentation

Let $l(\hat{a}) = \frac{1}{2}\|y - V\hat{a}\|_2^2$ and suppose that an upper bound L of $\|V^{\mathrm{T}}V\|$ is obtained. The basic idea of proximal first-order methods is to replace l with its linear approximation around a point w_t specific to iteration t. This leads to the computation of a proximity operator, and specifically $u_t = (\beta_t, \lambda_t) \leftarrow$

$$\arg\min\{\frac{L}{2}\left\|(\beta, \lambda) - (w_t - \frac{1}{L}\nabla l(w_t))\right\|_2^2$$

$+ \rho\Gamma(\beta, \lambda) : \beta \in {}^p, \lambda \in \Lambda\}$. And then, the

point w_t is updated rely on the current and

previous estimates of the solution u_t, u_{t-1}, . . .

and the process repeats. Our main algorithm for ontology sparse vector computing is sated as follows:

Ontology Sparse Vector Learning Algorithm:

u_1, w_1 $\leftarrow$ any feasible values

for t=1,2,. . . do

Calculate a fixed point $\hat{v}^{(t)}$ of H_t

$$u_{t+1} \leftarrow w_t \square \frac{1}{L}\nabla l(w_t) - \frac{c}{L}B^{\mathrm{T}}\hat{v}^{(t)}$$

$$w_{t+1} \leftarrow \pi_{t+1}u_{t+1} - (\pi_{t+1} - 1)u_t$$

end for

Specifically, let θ_t be the parameter defined

by $\theta_1 = 1$ and $\frac{1 - \theta_{t+1}}{\theta_{t+1}^2} = \frac{1}{\theta_t^2}$, then the optimal

update is $w_{t+1} \leftarrow u_{t+1} + \theta_{t+1}(\frac{1}{\theta_t} - 1)(u_{t+1} - u_t)$.

By computing the proximity operator of $\frac{\varphi}{L} \circ B$,

we rephrased the algorithm using the sequence

$\pi_t = 1 - \theta_t + \sqrt{1 - \theta_t} = 1 - \theta_t + \frac{\theta_t}{\theta_{t-1}}$. For

each iteration, using Picard-Opial process, H_t

is denoted as

$$H_t(x) = (I - \text{prox}_{\frac{\varphi}{c}})((I - \frac{c}{L}BB^{\mathrm{T}})^v$$

$$- \frac{1}{L}B(\nabla l(w_t) - Lw_t)).$$

Also, at stage s of the proximity calculating is

$v_{s+1} = kv_s + (1 - \kappa)H_t(v_s)$. Furthermore, the

fixed point process associated with the

assignment of u are equivalent to $u_{t+1} \leftarrow$

$\text{prox}_{\frac{\varphi}{L} \circ B}(w_t - \frac{1}{L}\nabla l(w_t))$. Let T be the total

times of iteration. Then the computation

complexity of above algorithm is $O(\frac{1}{T^2})$.

4. SIMULATION STUDIES

In this section, we designed two simulation experiments which are related to ontology

similarity measure and ontology mapping, respectively. Note that after getting the sparse vector $\hat{a}$, the ontology function then is derived

by $f_\beta(v) = \sum_{i=1}^{p} v_i \beta_i$ such that the noise term

δ is ignored.

4.1 Experiment on Plant Data

In this subsection, we use "PO" ontology O_1, constructed in http: //www.plantontology.org. (Fig. 1 shows the basic structure of O_1), to check the efficiency of our new algorithm in ontology similarity measuring. Similarly, the *P@N* standard [23] is used for this experiment. Moreover, the ontology methods in Wang et. al., [11], Huang et. al., [12] and Gao and Liang [13]

are employed to the "PO" ontology. We calculate the accuracy by the three algorithms, and at last compare the results with the new algorithm. Part of the data can be referred to Table 1.
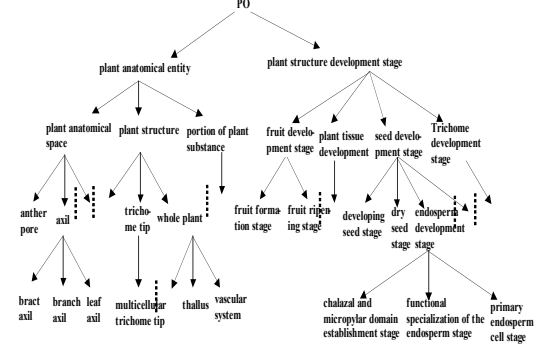

Fig. 1. "PO" ontology O_1

Table 1. The experiment data for ontology similairty measure

	P@3 average precision ratio	*P@5* average precision ratio	*P@10* average precision ratio
Our Algorithm	49.04%	57.85%	71.01%
Algorithm in [11]	45.49%	51.17%	58.59%
Algorithm in [12]	42.82%	48.49%	56.32%
Algorithm in [13]	48.31%	56.35%	68.71%

While *N*= 3, 5, or 10, the precision ratio which we get from our algorithm is higher than that determined by algorithms proposed in Wang et al., [11], Huang et al., [12] and Gao and Liang [13]. In particular, such precision ratios are increasing clearly as *N* increases. Thus, our algorithm is superior to the method presented by Wang et al., [11], Huang et al., [12] and Gao and Liang [13].

4.2 Experiment on University Data

"University" ontologies O_2 and O_3, whose structures are presented in Fig. 2 and Fig. 3 separately, are used for our last experiment. This experiment tends to determine ontology mapping between O_2 and O_3 by means of our ontology algorithm. *P@N* criterion is used to measure the equality of the experiment as well. Ontology algorithms in Wang et. al., [11], Huang et. al., [12] and Gao and Liang [13] are chosen to be applied to university ontologies as comparisons with the results obtained by our algorithm. In the end, through the comparisons of the precision ratio obtained by our algorithm and the four methods, we get several experiment results which
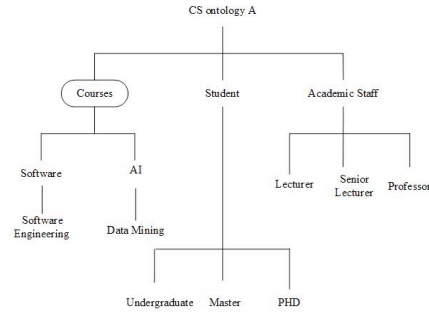
can be referred to Table 2.
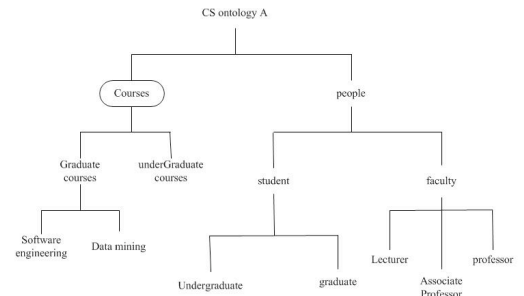

Fig. 2. "University" Ontology O_2

Fig. 3. "University" Ontology O_3.

Table 2. The experiment data for ontology mapping

	*P@*1 average precision ratio	*P@*3 average precision ratio	*P@*5 average precision ratio
Our Algorithm	57.14%	71.43%	84.29%
Algorithm in [11]	50.00%	59.52%	68.57%
Algorithm in [12]	42.86%	52.38%	60.71%
Algorithm in [13]	57.14%	64.29%	65.00%

The experiment results in Table 2 show that, our algorithm turns out to be more efficient than algorithms proposed in Wang et al., [11], Huang et al., [12] and Gao and Liang [13] especially in the situation where *N* is sufficiently large.

5. CONCLUSIONS

In this paper, the new ontology framework and its optimal technologies are presented for ontology sparse vector computation. The new iterative computation algorithm is based on proximity operator computation, Picard-Opial process and accelerated first-order tricks. At last, simulation data shows that our new algorithm has high efficiency in plant science and university ontologies. The ontology sparse algorithm raised in our paper illustrates the promising application prospects for multiple disciplines.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

REFERENCES

[1]   J. M. Przydzial, B. Bhatarai, and A. Koleti, GPCR ontology: development and application of a *G* protein-coupled receptor pharmacology knowledge framework. *Bioinformatics*, 29(24), 3211-3219, 2013.

[2]   S. Koehler, S. C. Doelken, and C. J. Mungall, The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Research*, 42(D1), 966-974,2014

[3]   M. Ivanovic, and Z. Budimac, An overview of ontologies and data resources in medical domains. *Expert Systerms and Applications*, 41(11), 5158-15166,2014.

[4]   A. Hristoskova, V. Sakkalis, and G. Zacharioudakis, Ontology-driven monitoring of patient's vital signs enabling personalized medical detection and alert. *Sensors*, 14(1), 1598-1628,2014.

[5]   M. A. Kabir, J. Han, and J. Yu, User-centric social context information management: an ontology-based approach and platform. *Personal and Ubiquitous Computing*, 18(5), 1061-1083,2014.

[6]   Y. L. Ma, L. Liu, K. Lu, B. H. Jin, and X. J. Liu, A graph derivation based approach for measuring and comparing structural semantics of ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 26(5), 1039-1052,2014.

[7]   Z. Li, H. S. Guo, Y. S. Yuan, and L. B. Sun, Ontology representation of online shopping customers knowledge in enterprise information. *Applied Mechanics and Materials*, ,483,603-606,2014.

[8]   R. Santodomingo, S. Rohjans, M. Uslar, J. A. Rodriguez-Mondejar, and M.A. Sanz-Bobi, Ontology matching system for future energy smart grids. *Engineering Applications of Artificial Intelligence*, 32, 242-257,2014.

[9]   T. Pizzuti, G. Mirabelli, M. A. Sanz-Bobi, and F. Gomez-Gonzalez, Food Track & Trace ontology for helping the food traceability control. *Journal of Food Engineering*,, 120(1), 17-30,2014.

[10]   N. Lasierra, A. Alesanco, and J. Garcia, Designing an architecture for monitoring patients at home: Ontologies and web services for clinical and technical management integration. *IEEE Journal of Biomedical and Health Informatics*, 18(3), 896-906,2014.

[11]   Y. Wang, W. Gao, Y. Zhang, and Y. Gao, Ontology similarity computation use ranking learning Method. *The 3rd*

International Conference on Computational Intelligence and Industrial Application, Wuhan, China, 2010, 20-22.

[12]   X. Huang, T. Xu, W. Gao, and Z. Jia, Ontology similarity measure and ontology mapping via fast ranking method. *International Journal of Applied Physics and Mathematic*, 1, 54-59,2011.

[13]   W. Gao, and L. Liang. Ontology similarity measure by optimizing NDCG measure and application in physics education. *Future Communication, Computing, Control and Management,* 142, 415-421,2011.

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

[15]   X. Huang, T. Xu, W. Gao, and S. Gong, 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.

[16]   W. Gao, Y. Gao, and L. Liang, Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping. *Journal of Chemical and Pharmaceutical Research*, 5(9), 592-598,2013

[17]   W. Gao, and L. Shi, Ontology similarity measure algorithm with operational cost and application in biology science. *BioTechnology: An Indian Journal,*, 8(11), 1572-1577,2013.

[18]   W. Gao, and T. W. Xu, Ontology similarity measuring and ontology mapping algorithm based on MEE criterion. *Energy Education Science and Technology Part A: Energy Science and Research*,32(5), 3793-3806,2014

[19]   W. Gao, Y. Gao, and Y. Zhang, Strong and weak stability of k-partite ranking algorithm. *Information,* 15(11A), 4585-4590,2012.

[20]   W. Gao, and T. W. Xu, Stability analysis of learning algorithms for ontology similarity computation. *Abstract and AppliedAnalysis,*2013,9pages,
http://dx.doi.org/10.1155/2013/174802.

[21]   W. Gao, and L. Zhu, Gradient learning algorithms for ontology computing. *Computational Intelligence and Neuroscience,*2014,12pages,
http://dx.doi.org/10.1155/2014/438291.

[22]   W. Gao, L. Yan, and L. Liang, 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 (3/4), 226-231,2014.

[23]   N. Craswell, and D. Hawking, Overview of the TREC 2003 web track. *Proceeding of the Twelfth Text Retrieval Conference,* Gaithersburg, Maryland, NIST Special Publication, 2003, 78-92.