



## Algorithm of Ontology Similarity Measure Based on Similarity Kernel Learning

Linli Zhu<sup>1</sup>, Xiaozhong Min<sup>1</sup>, Wei Gao<sup>2</sup>

<sup>1</sup> Jiangsu University of Technology, School of Computer Engineering, Changzhou, Jiangsu, China

<sup>2</sup> Yunnan Normal University, School of Information, Kunming, Yunnan, China

### ABSTRACT

Ontology, as a structured conceptual model of knowledge representation and storage, has widely been used in biomedical and pharmaceutical research. The nature of the ontology application is to get the similarity between ontology vertices, and thus reveal the similarity of their corresponding concepts and intrinsic relationships. The similarity for all pairs of vertices forms a similarity matrix of ontology, and the aim of ontology algorithm is to obtain the optimal similarity matrix. In this paper, we propose a new algorithm to get a similarity matrix in terms of learning the optimal similarity kernel function for ontology. The simulation experiment is designed for biology "Go" ontology application, and the result data reveal that new algorithm is efficient for such task.

### Indexing terms/Keywords

Ontology; Similarity measure; Centering kernel; Kernel alignment; Positive Semi-definite Matrix

### Academic Discipline And Sub-Disciplines

Computer Science and Technology; Machine Learning

# Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol. 14, No. 12

[www.ijctonline.com](http://www.ijctonline.com), [editorijctonline@gmail.com](mailto:editorijctonline@gmail.com)



## INTRODUCTION

In computer science, ontology is a model for knowledge storing and representation, and has been applied in image retrieval, knowledge management, information retrieval search extension, information systems, collaboration, and intelligent information integration. In recent years, as an effective concept semantic model, ontology has been widely employed in medical science, pharmacology science and biology science (for instance, see [1-3]).

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 simple undirected graph correspond to ontology  $O$ , the goal of ontology application is similarity measuring between ontology vertices, and the aim of similarity measure is to approach a similarity function which maps each pair of vertices to a non-negative real number.

In biology and pharmacy science, we use a vector to represent the information for each gene or disease. By studying the similarity of corresponding information between the genes, or the similarity of corresponding information between gene and diseases, we study the linkages between different genes, and what types of genes cause the certain disease.

There are some effective technologies for getting efficient ontology similarity measure algorithm. [4] first proposed that ranking method can be employed in ontology similarity calculation. [5] raised fast ontology algorithm in order to reduce the time complexity of the algorithm. [6] argued that ontology function can be given by optimizing NDCG[7] obtained the ontology function in terms of the regression measure, and applied such idea in physics education. approach. In [8] the proposal was getting ontology function based on half transductive ranking. [9] explored the learning theory approach for ontology similarity computation in a setting when the ontology graph is a tree. In view of 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  $1 \leq a < b \leq k$ ) 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. [10] proposed new criterion for multi-dividing ontology algorithm from AUC standpoint, which was designed to avoid the choice of loss function. [11] presented new algorithms for ontology similarity measurement and ontology mapping by virtue of harmonic analysis and diffusion regularization on hypergraph. Recently, [12] proposed new algorithms for ontology similarity measurement such that the new computational models consider operational cost in the real implement.

Several papers have contributed to the theoretical analysis for different ontology settings. [13] have studied the uniform stability of multi-dividing ontology algorithm and gave the generalization bounds for stable multi-dividing ontology algorithms. [14] researched the strong and weak stability of multi-dividing ontology algorithm. [15] learned some characteristics for such ontology algorithm. [16] 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. [17] presented the characteristics of best ontology score function among piece constant ontology score functions. [18] investigated the upper bound and the lower bound minimax learning rate which is obtained based on low noise assumptions. [19] and [20] presented an approach of piecewise constant function approximation for AUC criterion multi-dividing ontology algorithms.

Since the similarity of all pairs of vertices forms a similarity matrix, the aim of ontology application boils down to derive the optimal similarity matrix. In this paper, we focus on the ontology algorithm to deduce the best similarity matrix by learning the similarity kernel. The organization of this paper is as follows: the ontology problem and notations for basic setting are showed in Section 2; the computing model and basic ideals are stated in Section 3; in Section 4, an simulation experiment is designed to test the efficiency of the algorithm, and the data result reveal that our algorithm has high precision ratio for biology application.

## Setting and Notation

For each  $v \in V$ , we use a vector to express all the information of  $v$  in ontology graph. The elements in  $V$  are drawn independently and randomly according to some unknown distribution  $D$ . Without causing confusion, we use  $v$  to represent the vertex and the information vector corresponds to  $v$ . Given a training set  $S = \{v_1, \dots, v_m\} \subset V$  with size  $m$ , the aim of ontology learning algorithms is to get an optimal similarity function  $Sim^*: V \times V \rightarrow \mathbb{R}^+$ . Since the positive kernel function  $K: V \times V \rightarrow \mathbb{R}^+$  also return a non-negative real number for each pair of ontology vertices, we consider similarity kernel instead of similarity function. In this article, we determine the similarity between vertices via learning optimal similarity kernel function.

Recall that the feature mapping  $\Phi: V \rightarrow H$  is consisted of replacing it by  $\Phi - E_v[\Phi]$ , where  $E_v$  is the expected value of  $\Phi$  and vertex  $v$  is drawn from the distribution  $D$ . Kernel function  $K: V \times V \rightarrow \mathbb{R}^+$  is a positive definite symmetric (PDS) function which consists of information of feature mapping  $\Phi$  associated to  $K$ . For each pair of  $v, v' \in V$ , the centered kernel  $K_c$  associated to  $K$  is expressed by

$$K_c(v, v') = (\Phi(v) - E_v[\Phi])^T (\Phi(v') - E_{v'}[\Phi])$$

$$= K(v, v') - E_v[K(v, v')] - E_{v'}[K(v, v')] - E_{v, v'}[K(v, v')].$$

The definition doesn't rely on the selecting of the feature mapping associated to  $K$ . Note that  $K_c$  is also a PDS kernel since  $K_c(v, v')$  is regarded as an inner product. For any centered kernel  $K_c$ , we have  $E_{v, v'}[K_c(v, v')] = 0$  which implies that centering the feature mapping is just centering the kernel function.

Let  $\bar{\Phi} = \frac{1}{m} \sum_{i=1}^m \Phi(v_i)$ , then a feature vector  $\Phi(v_i)$  with  $i \in [1, m]$  for training set  $S$  be centered by replacing it with

$\Phi(v_i) - \bar{\Phi}$  and the kernel matrix  $K$  associated to the training set  $S$  and  $K$  be centered by replacing it with  $K_c$  denoted for any  $i, j \in [1, m]$  as

$$[K_c]_{ij} = K_{ij} - \frac{1}{m} \sum_{i=1}^m K_{ij} - \frac{1}{m} \sum_{j=1}^m K_{ij} - \frac{1}{m^2} \sum_{i, j=1}^m K_{ij}.$$

Set  $\Phi = [\Phi(v_1), \dots, \Phi(v_m)]^T$  and  $\bar{\Phi} = [\bar{\Phi}, \dots, \bar{\Phi}]^T$ . We get  $K_c = (\Phi - \bar{\Phi})(\Phi - \bar{\Phi})^T$ , which implies that  $K_c$  is also a positive semi-definite (PSD) matrix. Analogously, we obtain  $\frac{1}{m^2} \sum_{i, j=1}^m [K_c]_{ij} = 0$ .

Let  $K$  and  $K'$  be two kernel functions defined on  $V \times V$  with  $0 < E[K_c^2] < +\infty$  and  $0 < E[K_c'^2] < +\infty$ . Hence, the alignment between  $K$  and  $K'$  is denoted as  $\rho(K, K') = \frac{E[K_c K_c']}{\sqrt{E[K_c^2] E[K_c'^2]}}$ .

In order to abbreviate the notations, we usually skip the variables over which an expectation is taken. Since  $|E[K_c K_c']| \leq \sqrt{E[K_c^2] E[K_c'^2]}$  in terms of Cauchy-Schwarz inequality, we infer  $\rho(K, K') \in [-1, 1]$ . Let  $\langle \cdot, \cdot \rangle_F$  be the Frobenius product and  $\|\cdot\|_F$  be the Frobenius norm. We derive  $E[QQ'] \geq 0$  for any two PDS kernels  $Q$  and  $Q'$ .

Let  $K \in \square^{m \times m}$  and  $K' \in \square^{m \times m}$  be two kernel matrices satisfy  $\|K_c\|_F \neq 0$  and  $\|K_c'\|_F \neq 0$  respectively. Thus, the

alignment between  $K$  and  $K'$  is denoted as  $\hat{\rho}(K, K') = \frac{\langle K_c, K_c' \rangle_F}{\|K_c\|_F \|K_c'\|_F}$ .

By virtue of the Cauchy-Schwarz inequality, we yield  $\hat{\rho}(K, K') \in [-1, 1]$ . More accurately, we verify that  $\hat{\rho}(K, K') \geq 0$  in terms of the fact that the Frobenius product of any two positive semi-definite matrices  $K$  and  $K'$  is non-negative. For such matrices, there exist matrices  $U_1$  and  $U_2$  satisfy  $K = U_1 U_1^T$  and  $K' = U_2 U_2^T$  which follow from  $\langle K, K' \rangle_F = \text{Tr}(U_1 U_1^T U_2 U_2^T) = \text{Tr}((U_1^T U_2)^T (U_1^T U_2)) \geq 0$ .

### Similarity Kernel Learning for Ontology Algorithm

In this section, we consider our ontology algorithms via learning kernels in the form of linear combinations of  $p$  base kernels  $K_k$ , where  $k \in [1, p]$ . In such situation, the final hypothesis learned belongs to the reproducing kernel Hilbert space

(RKHS) associated with a kernel  $K_\mu = \sum_{k=1}^p \mu_k K_k$ . Notice that in order to guarantee  $K$  is a PDS kernel, the mixture

weights are choice subject to the condition  $\mu \geq 0$  and  $\|\mu\| = \Lambda > 0$  is a regularization parameter. Our algorithm is to determine the mixture weights  $\mu$ .

W.l.o.g., we assume that the centered base kernel matrices  $K_{k_c}$  are independent. Hence, we infer  $\|K_{\mu_c}\|_F > 0$  for any  $\mu$  and ensure that  $\hat{\rho}(K_\mu, \mathbf{y} \mathbf{y}^T)$  is well defined. By the fashion of the properties of centering, we deduce  $\langle K_{\mu_c}, K_{Y_c} \rangle_F = \langle K_{\mu_c}, K_{Y_c} \rangle_F$ . Let  $M = \{\mu : \|\mu\|_2 = 1\}$ . According to the fact that  $\|K_{Y_c}\|_F$  doesn't rely on  $\mu$ , alignment

maximization can be expressed as the following optimization problem:

$$\max_{\mu \in M} \hat{\rho}(K_{\mu}, yy^T) = \max_{\mu \in M} \frac{\langle K_{\mu_c}, yy^T \rangle_F}{\|K_{\mu_c}\|_F}. \quad (1)$$

A similar setting can be designed by virtue of 2-norm. Let  $K_{\mu_c} = U_m K_{\mu} U_m$  with  $U_m = I - 11^T / m$ . We obtain  $K_{\mu_c} = \sum_{k=1}^p \mu_k U_m K_k U_m = \sum_{k=1}^p \mu_k K_{k_c}$ . Let  $\mathbf{a} = (\langle K_{1_c}, yy^T \rangle_F, \dots, \langle K_{p_c}, yy^T \rangle_F)^T$  and  $\mathbf{M}$  be the matrix expressed by  $M_{kl} = \langle K_{k_c}, K_{l_c} \rangle_F$  for each  $k, l \in [1, p]$ . Since we assumed that the base kernels are independent, matrix is invertible. In terms of the non-negativity of the Frobenius product of PSD matrices, we ensure that the entries of  $\mathbf{a}$  and  $\mathbf{M}$  are all non-negative. Furthermore,  $\mathbf{M}$  is a symmetric PSD matrix since for each vector  $V = (v_1, \dots, v_m) \in \mathbb{R}^m$ , we have

$$\begin{aligned} V^T M V &= \sum_{k,l=1}^m v_k v_l \text{Tr}[K_{k_c} K_{l_c}] = \text{Tr}[\sum_{k,l=1}^m v_k v_l K_{k_c} K_{l_c}] \\ &= \text{Tr}[(\sum_{k=1}^m v_k K_{k_c})(\sum_{l=1}^m v_l K_{l_c})] \\ &= \left\| \sum_{k=1}^m v_k K_{k_c} \right\|_F^2 \geq 0. \end{aligned}$$

Thus, the solution  $\mu^*$  of the optimization problem (1) is determined by  $\mu^* = \frac{M^{-1} \mathbf{a}}{\|M^{-1} \mathbf{a}\|}$ . The alignment maximization problem with the set  $M := \{\|\mu\|_2 = 1 \wedge \mu \geq 0\}$  can be expressed as  $\mu^* = \arg \max_{\mu \in M} \frac{\mu^T \mathbf{a} \mathbf{a}^T \mu}{\mu^T \mathbf{M} \mu}$ .

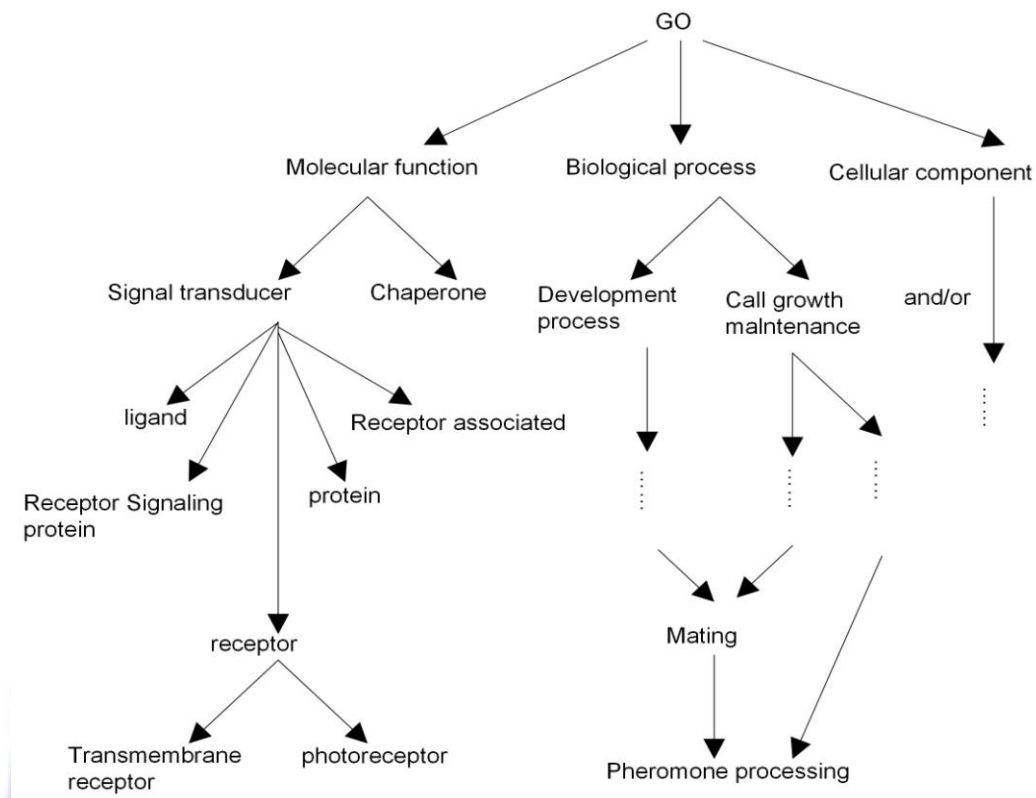
### Simulation Experiment

In this section, we design the following simulation experiment concerning ontology similarity measures. We use biology ontology  $O_1$  which was constructed in <http://www.geneontology.org>. Fig. 1 shows the basic structure of  $O_1$ .  $P@N$  (Precision Ratio, see [21] for more detail) is used to quantify the quality of our experiment. First, we obtain the closest  $N$  concepts for every vertex on the ‘‘GO’’ ontology graph by biology expert, and then we obtain the first  $N$  concepts for every vertex on ontology graph by the algorithm and compute the precision ratio. At the same time, we apply ontology method in [8], and [22] to the ‘‘GO’’ ontology. We calculate the accuracy by these two algorithms and compare the result to algorithm proposed in our paper, part of the data refer to Tab.1.

**Tab. 1 The experiment results of ontology similarity measure**

	$P@3$ average precision ratio	$P@5$ average precision ratio	$P@10$ average precision ratio	$P@20$ average precision ratio
<b>Algorithm in our paper</b>	57.43%	66.34%	77.45%	88.21%
<b>Algorithm in [8]</b>	47.73%	55.62%	69.93%	76.82%
<b>Algorithm in [22]</b>	52.37%	60.62%	72.96%	78.64%





**Fig. 1: “GO” Ontology  $O_1$**

By the experiment data presented in Tab. 1, we arrived at the conclusion that our algorithm is more efficiently than algorithms raised in [8], and [22] especially when  $N$  is large enough. Therefore, this new ontology similarity algorithm by virtue of centered kernel learning has high efficiency.

## Summary

Ontology, as a data representation model, has been widely used in medical science, pharmacology science and biology science. And, it is proved to be highly efficient. In our paper, we apply the tricks of kernel learning to design the new similarity measure algorithms for ontology. The new algorithms have high quality according to the simulation data presented in the former section. This technology contributes to the state of art for ontology in biology and pharmacology applications.

## Acknowledgements

We thank the reviewers for their constructive comments in improving the quality of this paper. We also would like to thank the anonymous referees for providing us with constructive comments and suggestions. We wish to acknowledge the National Natural Science Foundation of China (61142007) and the Natural Science Foundation of Jiangsu University of Technology in China (KYY14013).

## References

- [1] Hu, B., Dasmahapatra, S., Lewis, P. and Shadbolt, N. ‘Ontology-based medical image annotation with description logics’, In 15th IEEE International Conference on Tools with Artificial Intelligence, California, Sacramento, USA, (2003) :77-82.
- [2] Lambrix, P., Edberg, A., ‘Evaluation of ontology tools in bioinformatics’, Paci Symposium on Biocomputing, New York: IEEE Computer Society Press, (2003):529-600.
- [3] Mork, P., Bernstein, P., ‘Adapting a generic match algorithm to align ontologies of human anatomy’ In 20th International conf on Data engineering. New York: IEEE Computer Society Press, (2004):787-790.
- [4] Wang, Y., Gao W., Zhang, Y. and Gao Y. ‘Ontology similarity computation use ranking learning Method’, The 3rd International Conference on Computational Intelligence and Industrial Application, Wuhan, China, (2010):20-22.
- [5] Huang, X., Xu, T., Gao, W. and Jia, Z. ‘Ontology similarity measure and ontology mapping via fast ranking method’, International Journal of Applied Physics and Mathematics, Vol. 1, No. 1, (2011): 54-59.
- [6] Gao, W. and Liang, L. ‘Ontology similarity measure by optimizing NDCG measure and application in physics education’, Future Communication, Computing, Control and Management, Vol. 142, (2011) :415–421.



- [7]Gao, Y. and Gao, W. 'Ontology similarity measure and ontology mapping via learning optimization similarity function', International Journal of Machine Learning and Computing, Vol. 2, No 2. (2012):107-112.
- [8]Huang, X., Xu, T., Gao, W. and 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.
- [9]Lan, M., Xu, J. and Gao, W. (2012) 'Ontology similarity computation using k-partite ranking method', Journal of Computer Applications, Vol. 32, No. 4, (2011):1094-1096.
- [10]Zhu, L. and Gao, W. 'Theoretical Analysis for New Multi-dividing Ontology Algorithm Based on AUC Criterion', Manuscript. (2015)
- [11]Gao, W., Gao Y. and Liang L. 'Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping', Journal of Chemical and Pharmaceutical Research, 9 (2013):592-598.
- [12]Gao, W. and Shi, L. 'Ontology similarity measure algorithm with operational cost and application in biology science', BioTechnology : An Indian Journal, In press. (2013)
- [13]Gao, W. and Xu, T. 'Stability analysis of learning algorithms for ontology similarity computation', Abstract and Applied Analysis, Vol .2013 (2013), Article ID 174802, 9 pages. (2013)
- [14]Gao, W., Gao, Y. and Zhang, Y. 'Strong and weak stability of k-partite ranking algorithm', Information, Vol. 15, No 11A, (2012):4585-4590.
- [15]Gao, W. and Xu, T. 'Characteristics of optimal function for ontology similarity measure via multi-dividing', Journal of networks, Vol.7, No 8, (2012):1251-1259.
- [16] Gao, W., Xu, T., Gan, J. and Zhou, J. 'Linear Statistical Analysis of Multi-dividing Ontology Algorithm', Journal of Information and Computational Science, In press. (2014)
- [17]Gao, Y., Gao, W. and Liang, L. 'Statistical Characteristics for Multi-dividing Ontology Algorithm in AUC Criterion Setting', Manuscript. (2014)
- [18]Gao, W., Gao, Y. Zhang, Y. and Liang, L. 'Minimax Learning Rate for Multi-dividing Ontology Algorithm', Manuscript. (2014)
- [19]Gao, W., Yan, L. and Liang, L. 'Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (I)', International Journal of Computer Applications in Technology, In press. (2013)
- [20]Yan, L., Gao, W. and Li, J. 'Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (II) ', Journal of Applied Science, 16 (2013): 3257-3262.
- [21]Craswell, N. and D. Hawking, 'Overview of the TREC 2003 web track', Proceedings of the Twelfth Text Retrieval Conference. Gaithersburg, Maryland, NIST Special Publication, (2003):78-92.
- [22]Gao, W. and M. Lan, 'Ontology mapping algorithm based on ranking learning method', Microelectronics and Computer, 9(2011): 59-61.

### Author' biography with Photo



Linli,Zhu, male, was born in the city of Badong, Hubei Province,China on Sep. 20, 1975. He got Master degrees on computer software and theory from Yunnan normal university in 2007. Now, he acts as associate professor in the department of computer engineering, Jiangsu University of Technology. As a researcher in computer science, his interests are covered two disciplines: computer network and artificial intelligence.