Similarity matrix learning for ontology application

Jianzhang Wu*

School of Computer Science and Engineering,
Southeast University,
Nanjing 210096, China
Email: jzwu@njnet.edu.cn
*Corresponding author

Xiao Yu

School of Continuing Education,
Southeast University,
Nanjing 210096, China
Email: pp_xyu@seu.edu.cn

Wei Gao

School of Information Science and Technology,
Yunnan Normal University,
Kunming 650500, China
Email: gaowei@ynnu.edu.cn

Abstract: In information retrieval, ontology is used to search the information which has highly semantic similarity of the original query concept, and return the results to the user. Ontology mapping is used to create the relationship between different ontologies, and the essence of which is similarity computation. In this article, we present new algorithms for ontology similarity measure and ontology mapping by determining the similarity matrix of ontology. The optimisation strategy and iterative procedure are designed in terms of metric distance learning tricks. The simulation experimental results show that the proposed new algorithms have high accuracy and efficiency on ontology similarity measure and ontology mapping in biology, physics applications, plant science and humanoid robotics.

Keywords: knowledge representation; ontology; similarity measure; ontology mapping; similarity matrix.

Reference to this paper should be made as follows: Wu, J., Yu, X. and Gao, W. (2016) 'Similarity matrix learning for ontology application', *Int. J. Information Technology and Management*, Vol. 15, No. 1, pp.1–13.

Biographical notes: Jianzhang Wu received his BSc in 1995 from the Department of Automation, Master in 2002 in the fields of computer application from Southeast University. He began to teach and research in the field of computer network and its application in early 2000s, and major research topics covers information search engine, machine learning, cloud computing and the behaviour of larger-scale network. He has published more than ten papers and one textbook since 1997, and taken two research projects and received two academic awards from central and provincial government.

Xiao Yu received her BSc and Master in 1995 and 2002 in the fields of computer application from Southeast University. She began to research in the field of computer network and its application in early 2000s, and major research topics covers web technology, information search engine and cloud computing. She has published more than eight papers since 1997, and received one academic award from central and provincial government.

Wei Gao received his Bachelor degree in Computer Science from Zhejiang Industrial University in 2004 and Mathematics Education from College of Zhejiang education in 2006. Then, he received his Master degree from Department of Computer Science and Information Technology, Yunnan Normal University in 2009. In 2012, he received his PhD in Department of Mathematics, Soochow University, China. Currently, he acts as a 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.

1 Introduction

As a knowledge representation and conceptual shared model, ontology has been applied in image retrieval, knowledge management and information retrieval search extension. Acting as an effective concept semantic model, ontology also employed in disciplines beyond computer science, such as social science (for instance, see Bouzeghoub and Elbyed, 2006), biology science (for instance, see Hu et al., 2003) and geography science (for instance, see Fonseca et al., 2001).

The ontology model is actually a graph $G = (V, E)$, each vertex v in an ontology graph G represents a concept and each edge $e = v_iv_j$ on an ontology graph G represents a relationship between concepts v_i and v_j. The target of ontology similarity measure is to find a similarity function Sim: $V \times V \rightarrow \mathbb{R}^+ \cup \{0\}$ which maps each pair of vertices to a real number. The aim of ontology mapping is to bridge the link between two or more ontologies. Let G_1 and G_2 be two ontology graphs corresponding ontology O_1 and O_2 respectively. For each $v \in G_1$, find a set $S_v \subseteq V(G_2)$ where the concepts correspond to vertices in S_v are semantic close to the concept correspond to v. One method to get such mapping is, for each $v \in G_1$, computing the similarity $S(v, v_j)$ where $v_j \in V(G_2)$ and choose a parameter $0 < M < 1$. Then S_v is a collection such that the element in S_v satisfies $S(v, v_j) \geq M$. In this point of view, the essence of ontology mapping is to obtain a similarity function S and select a suitable parameter M. In our article, we focus on the technologies to yield a similarity matrix by virtue of distance learning.

For ontology similarity measure and ontology mapping, there have several effective learning tricks. Wang et al. (2010) proposed to learn a score function which mapping each vertex to a real number, and the similarity between two vertices can be measured according to the difference of real number they correspond to. Huang et al. (2011a) presented fast ontology algorithm for calculating the ontology similarity in a short time. Gao and Liang (2011) raised that the optimal ontology function can be determined by optimising NDCG measure, and applied such idea in physics education. Gao and Gao (2012) deduced the ontology function using the regression approach. Huang et al. (2011b) obtained ontology similarity function based on half transductive learning.

Gao and Xu (2013) explored the learning theory approach for ontology similarity computation using k-partite ranking method. Zhu and Gao (2014) proposed new criterion for ontology computation from AUC and multi-dividing standpoint. Gao et al. (2013) presented new ontology mapping algorithm using harmonic analysis and diffusion regularisation on hypergraph. Very recently, Gao and Shi (2013) proposed new ontology similarity computation technology such that the new calculation model consider operational cost in the real implement. Several theoretical analysis of ontology algorithm cans refer to Gao et al. (2012), Gao and Xu (2012) and Yan et al. (2013).

In this paper, we determine the new ontology similarity computation and ontology mapping algorithms based on metric distance learning tricks. Using the optimisation algorithm, we determine the optimal matrix to compute the similarity of vertices. The experiments are designed to show the efficiency of the algorithms.

2 Method of similarity matrix learning

First, we use a vector to represent the information of each vertex in ontology graph. Let $\{v_1, \ldots, v_n\} \in \mathbb{R}^d$ be n vertices. Our aim in this paper is to seek a positive definite similarity matrix A which parameterises the standard squared distance.

$$d_A(v_i, v_j) = (v_i - v_j)^T \boldsymbol{A}(v_i - v_j) \tag{1}$$

Suppose that we know the prior knowledge about intervertex distances. Consider relationships restricting the dissimilarity or similarity between pairs of ontology vertices. Two ontology vertices are similar if the distance between them is not larger than a given upper bound, i.e., $d_A(v_i, v_j) \le u$ for a relatively small value of parameter u. Similarly, two ontology vertices are dissimilar if $d_A(v_i, v_j) \ge l$ for sufficiently large parameter l.

A set of intervertex distance restrains are given as described above, and our problem is to learn a positive-definite similarity matrix $\boldsymbol{A}$ that parameterises the corresponding distance (1). Specifically, this learned similarity function is used to improve the accuracy of a k-nearest neighbour in ontology graph, or to incorporate semi-supervision into a distance-based learning algorithm. In many settings, we know the prior information about the distance function itself. If data is Gaussian, we parameterise the distance function in view of the inverse of the sample covariance. Hence, we regularise the similarity matrix $\boldsymbol{A}$ to be as close as possible to a given distance function, parameterised by $\boldsymbol{A}_0$. This implies that $\boldsymbol{A}_0$ is a given matrix with its elements $(\boldsymbol{A}_0)_{ij}$ determined by squared distance, and our optimal similarity matrix will as close as possible to $\boldsymbol{A}_0$.

We quantify the measure of closeness between optimal similarity matrix $\boldsymbol{A}$ and given distance matrix $\boldsymbol{A}_0$ by a natural information-theoretic exists a simple bijection between the set of equal mean multivariate Gaussian distributions with mean μ and the set of distances. For given distance parameterised by A, its corresponding multivariate Gaussian is expressed as

$$p(x, \boldsymbol{A}) = \frac{1}{Z}\exp\left\{-\frac{1}{2}d_A(v, \mu)\right\},$$

where A^{-1} is the covariance of the distribution and Z is a normalising constant. By virtue of such bijection, the distance between two distance functions parameterised by $\boldsymbol{A}_0$ and $\boldsymbol{A}$

is measured by the differential relative entropy between their corresponding multivariate Gaussians approach, i.e.,

$$KL\left(p\left(v;\boldsymbol{A}_0\right)\|p(v;\boldsymbol{A})\right)=\int p(v;\boldsymbol{A})\log\frac{p\left(v;\boldsymbol{A}_0\right)}{p(v;\boldsymbol{A})}dx \tag{2}$$

The distance (2) presents a well-founded measure of closeness between two distance functions. Based on (2), our distance metric learning problem for given pairs of similar vertices S and pairs of dissimilar vertices D can be determined as

$$\begin{aligned}&\min_{\boldsymbol{A}}\ KL\left(p(v;\boldsymbol{A}_0)\|p(v;\boldsymbol{A})\right)\\&\text{s.t.}\ \ d_A\left(v_i,v_j\right)\le u\ (i,j)\in S\\&\qquad d_A\left(v_i,v_j\right)\ge l\ (i,j)\in D\end{aligned} \tag{3}$$

In this paper, we use the technology of LogDet divergence which is a Bregman matrix divergence generated by the convex function $\phi(V) = -\log \det V$ defined over the cone of positive-definite matrices. For $n \times n$ matrices $\boldsymbol{A}$, $\boldsymbol{A}_0$, it equals

$$D_{ld}\left(\boldsymbol{A},\boldsymbol{A}_0\right)=\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{A}_0^{-1}\right)-\log\det\left(\boldsymbol{A}\boldsymbol{A}_0^{-1}\right)-n. \tag{4}$$

Note that the differential relative entropy between two multivariate Gaussians can be expressed as the convex combination of a distance between mean vectors and the LogDet divergence between the covariance matrices (for more detail, see Davis and Dhillon, 2006). Suppose the means of the Gaussians to be the same, we deduce

$$KL\left(p\left(v;\boldsymbol{A}_0\right)\|p(v;\boldsymbol{A})\right)=\frac{1}{2}D_{ld}\left(\boldsymbol{A}_0^{-1},\boldsymbol{A}^{-1}\right)=\frac{1}{2}D_{ld}\left(\boldsymbol{A},\boldsymbol{A}_0\right) \tag{5}$$

By Lehmann and Casella (2003), Stein's loss is the unique scale invariant loss-function such that the uniform minimum variance unbiased estimator is also a minimum risk equivariant estimator. In ontology metric learning, the scale invariance reveals that the divergence (4) remains invariant under any scaling of the feature space. In terms of

$$D_{ld}(\boldsymbol{S}^T\boldsymbol{A}\boldsymbol{S},\boldsymbol{S}^T\boldsymbol{B}\boldsymbol{S})=D_{ld}(\boldsymbol{A},\boldsymbol{B}), \tag{6}$$

the conclusion under any invertible linear transformation S can be further generalised to invariance. To represent the distance metric ontology learning problem (3), we exploit the equivalence in (5) and infer the following LogDet optimisation problem

$$\begin{aligned}&\min_{A\succ 0}\ D_{ld}\left(\boldsymbol{A},\boldsymbol{A}_0\right)\\&\text{s.t.}\ \ \operatorname{tr}\left(\boldsymbol{A}\left(v_i-v_j\right)\left(v_i-v_j\right)^T\right)\le u\ (i,j)\in S\\&\qquad \operatorname{tr}\left(\boldsymbol{A}\left(v_i-v_j\right)\left(v_i-v_j\right)^T\right)\ge l\ (i,j)\in D\end{aligned} \tag{7}$$

Note that the distance restrains on $d_A(v_i, v_j)$ become the above linear restrains on $\boldsymbol{A}$.

In certain situations, the feasible solution to (7) is not existed. To avoid such scenario happen, we incorporate slack variables into the formulation to ensure the existence of a feasible $\boldsymbol{A}$. Let $c(i, j)$ be the index of the $(i, j)^{\text{th}}$ constraint and ξ be a vector of slack variables, initialised to ξ_0. Its components equal u for similarity constraints and l for dissimilarity constraints. Then, we present the following optimisation problem instead of (7)

$$\begin{aligned} \min_{A \succ 0, \xi} \quad & D_{ld}\left(\boldsymbol{A}, \boldsymbol{A}_0\right) + \gamma D_{ld}\left(\operatorname{diag}(\xi), \operatorname{diag}(\xi_0)\right) \\ \text{s.t.} \quad & \operatorname{tr}\left(\boldsymbol{A}\left(v_i - v_j\right)\left(v_i - v_j\right)^T\right) \le \xi_{c(i,j)} \ (i,j) \in S \\ & \operatorname{tr}\left(\boldsymbol{A}\left(v_i - v_j\right)\left(v_i - v_j\right)^T\right) \ge \xi_{c(i,j)} (i,j) \in D \end{aligned} \tag{8}$$

Here, $D_{ld}(\boldsymbol{A}, \boldsymbol{A}_0)$ is used to measure the difference between $\boldsymbol{A}$ and $\boldsymbol{A}_0$; $D_{ld}(\text{diag}(\xi),\text{diag}(\xi_0))$ is employed to determine the gap between slack variables; γ is the parameter to control the tradeoff between satisfying the constraints and minimising $D_{ld}(\boldsymbol{A}, \boldsymbol{A}_0)$; $\boldsymbol{A} \succ 0$ denotes that $\boldsymbol{A}$ is a positive defined matrix; two restrictive conditions are used to control the distances of ontology vertices under matrix $\boldsymbol{A}$ according to the collection which the vertex pair (v_i, v_j) belong to.

To solve the optimisation problem (8), we use the extension methods of Kulis et al. (2006). The optimisation method which forms the basis for the algorithm repeatedly computes Bregman projections, i.e., projections of the current solution onto a single constraint. This projection is implemented via the update

$$\boldsymbol{A}_{t+1} = \boldsymbol{A}_t + \beta \boldsymbol{A}_t \left(v_i - v_j\right)\left(v_i - v_j\right)^T \boldsymbol{A}_t, \tag{9}$$

where v_i and v_j are the constrained data ontology vertices, and β is the projection parameter calculated by the algorithm. Each constraint projection has complexity $O(d^2)$, and hence a single iteration of looping via all constraints has complexity $O(cd^2)$. We emphasise that no eigen-decomposition is required in the algorithm. The resulting algorithm is presented in Algorithm 1. The inputs to the algorithm are the starting matrix A_0, the constraint data, and the slack parameter γ. If necessary, the projections can be measured efficiently over a factorisation W of the matrix, such that $\boldsymbol{A} = W^T\boldsymbol{W}$.

Algorithm 1

Input: V: input $d \times n$ matrix; S: set of similar pairs; D: set of dissimilar pairs; u, l: distance thresholds; A_0: input Mahalanobis matrix; γ: slack parameter; c: constraint index function

Output: A: output similarity matrix

$A \leftarrow A_0$, $\lambda_{ij} \leftarrow 0$, for any i and j.

$\xi_{c(i,j)} \leftarrow u$ for $(i, j) \in S$; otherwise $\xi_{c(i,j)} \leftarrow l$;

Repeat

pick a constraint $(i,j) \in S$ or $(i, j) \in D$.

$p \leftarrow (v_i, v_j)^T A_t(v_i, v_j)$, $\delta \leftarrow 1$ if $(i,j) \in S$, -1 otherwise

$$\alpha \leftarrow \min(\lambda_{ij}, \frac{\delta}{2}(\frac{1}{p} - \frac{\gamma}{\xi_{c(i,j)}})), \beta \leftarrow \frac{\delta\alpha}{1 - \delta\alpha p} \xi_{c(i,j)} \leftarrow \frac{\gamma \xi_{c(i,j)}}{\gamma + \delta\alpha\xi_{c(i,j)}}$$

$\lambda_{ij} \leftarrow \lambda_{ij} - \alpha$, $A \leftarrow A + \beta A(v_i, v_j)(v_i, v_j)^T A$

until convergence

Return ontology similarity matrix A

3 Some extension of ontology algorithm

In this section, we consider Kernelising metric ontology learning algorithm. Suppose that $A_0 = I$, i.e., the maximum entropy formulation that regularises to the baseline Euclidean distance. It is possible to Kernelise for other selections of A_0, but not presented. If $A_0 = \boldsymbol{I}$, the corresponding K_0 from the low-rank kernel ontology learning problem becomes $\boldsymbol{K}_0 = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}$, the Gram matrix of the inputs ontology data. If instead of an explicit representation V of our data vertices, we have as input a kernel function $\kappa(v_i, v_j) = \phi(v_i)^T\phi(v_i)$ with the associated kernel matrix $\boldsymbol{K}_0$ over the training ontology vertices, a natural question to ask is whether we can evaluate the learned metric on new ontology vertices in the kernel space. This requires the calculation of

$$\begin{aligned}
& d_A\left(\phi(v_i),\phi(v_j)\right)\frac{n!}{r!(n-r)!} \\
&= \left(\phi(v_i),\phi(v_j)\right)^T \boldsymbol{A}\left(\phi(v_i),\phi(v_j)\right) \\
&= \phi(v_i)^T \boldsymbol{A}\phi(v_i) - 2\phi(v_i)^T \boldsymbol{A}\phi(v_j) + \phi(v_j)^T \boldsymbol{A}\phi(v_j). \\
&= \phi(v_i)^T \boldsymbol{A}\phi(v_i)
\end{aligned}$$

Let $\tilde{\kappa}(v_i, v_j) = \phi(v_i)^T \boldsymbol{A}\phi(v_i)$ be new kernel function. The ability to generalise to unseen data of ontology vertices reduces to the ability to calculate $\tilde{\kappa}(v_i, v_j)$. Note that $\boldsymbol{A}$ can be regarded as an operator in a Hilbert space, and its size is just the dimensionality of $\varphi(v)$ which can potentially be infinite.

Although $\boldsymbol{A}$ cannot be explicitly computed, it is still possible to calculate $\tilde{\kappa}(v_i, v_j)$. Set $\boldsymbol{A}_0 = \boldsymbol{I}$, the learned $\boldsymbol{A}$ matrix can be recursively unrolled since it is of the form

$$\boldsymbol{A} = I + \sum_{i',j'} \sigma_{ij}\phi\left(v_i^{i'}\right)\phi\left(v_i^{j'}\right)^T$$

which followed by expanding equation (9) down to $\boldsymbol{I}$. The new kernel function is hence computed by

$$\tilde{\kappa}\left(v_i - v_j\right) = \kappa\left(v_i - v_j\right) + \sum_{i',j'} \sigma_{ij}\kappa\left(v_i^{i'}, v_j\right)\kappa\left(v_i^{j'}, v_j\right),$$

and is a function of the σ_{ij} coefficients and the original kernel function κ. While minimising $D_{ld}(\boldsymbol{K}, \boldsymbol{K}_0)$, the σ_{ij} coefficients can be updated without affecting the asymptotic running time of the algorithm; i.e., in terms of optimising the following low-rank kernel learning problem for K, the necessary coefficients σ_{ij} for evaluation of $\tilde{\kappa}(v_i, v_j)$ are yielded:

$$\begin{aligned}
& \min_{\boldsymbol{K}\succ 0} D_{ld}\left(\boldsymbol{K}, \boldsymbol{K}_0\right) \\
& \text{s.t. } K_{ii} + K_{jj} - 2K_{ij} \le u \ (i,j) \in S \\
& \qquad K_{ii} + K_{jj} - 2K_{ij} \ge l \ (i,j) \in D
\end{aligned}$$

This is to a trick for searching the nearest neighbour of a new ontology vertices in the kernel space under the learned metric with complexity $O(n^2)$.

4 Experiments

In this section, four simulation experiments relevance ontology similarity measure and ontology mapping are designed below. In order to adjacent to the setting of ontology algorithm, we use a vector with d dimension to express each vertex's information. Such vector contains the information of name, instance, attribute and structure of vertex. Here the instance of vertex refers to the set of its reachable vertex in the directed ontology graph.

4.1 Experiment on biology data

We use 'Go' ontology O_1 which was constructed in http://www.geneontology.org (Figure 1 shows the basic structure of O_1) for our experiment. *P@N* [precision ratio, see Craswell and Hawking (2003) for more detail] is used to measure the equality of the experiment. We first give the closest N concepts for every vertex on the ontology graph by expert, and then we obtain the first N concepts for every vertex on ontology graph by the algorithm and compute the precision ratio. Ontology algorithms in Huang et al. (2011a), Gao and Liang (2011) and Gao and Gao (2012) are employed to 'Go' ontology, and we compare the precision ratio which we get from four methods. Several experiment results refer to Table 1.

Figure 1 'Go' ontology

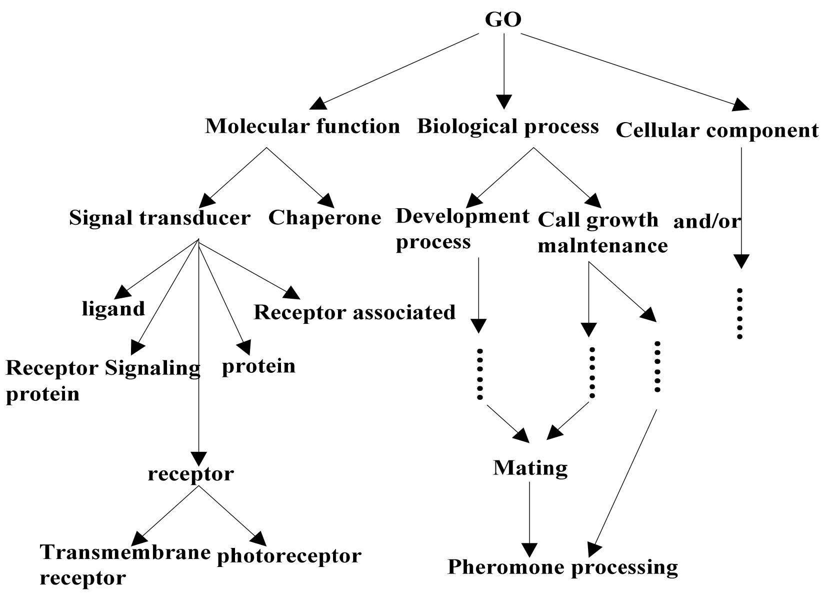
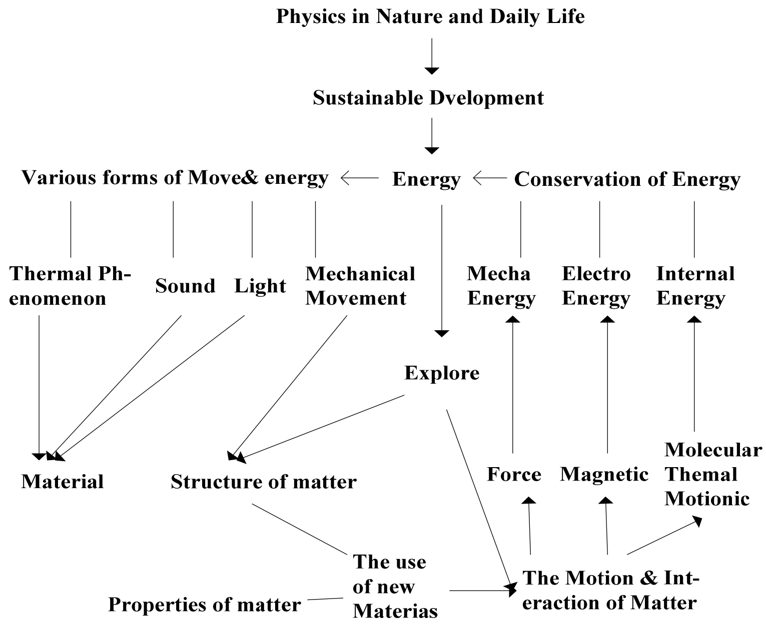


Table 1 The experiment data for ontology similairty measure

	P@3 average precision ratio	*P@5 average precision ratio*	*P@10 average precision ratio*	*P@20 average precision ratio*
Our algorithm	48.53%	56.91%	66.49%	78.83%
Algorithm in Huang et al. (2011a)	46.38%	53.48%	62.34%	74.59%
Algorithm in Gao and Liang (2011)	43.56%	49.38%	56.47%	71.94%
Algorithm in Gao and Gao (2012)	42.13%	51.83%	60.19%	72.39%

When $N = 3, 5, 10$ or 20, the precision ratio by virtue of our algorithm is higher than the precision ratio determined by algorithms proposed in Huang et al. (2011a), Gao and Liang (2011) and Gao and Gao (2012). In particular, when N increases, such precision ratios are increasing apparently. Therefore, the algorithm described in our paper is superior to the method proposed by Huang et al. (2011a), Gao and Liang (2011) and Gao and Gao (2012).

Figure 2 'Physical education' ontology O_2



4.2 *Experiment on physical education data*

We use physical education ontologies O_2 and O_3 (the structures of O_2 and O_3 are presented in Figures 2 and 3, respectively) for our second experiment. The goal of this experiment is determining the ontology mapping between O_2 and O_3 via similarity matrix which are deduced by Algorithm 1. *P@N* criterion is applied to measure the equality of the experiment. We first give the closest N concepts for each vertex on the ontology graph with the help of experts, and then we obtain the first N concepts for every vertex on ontology graph by the algorithm and compute the precision ratio. Also, ontology algorithms in Huang et al. (2011a), Gao and Liang (2011) and Gao et al. (2013) are employed to 'physical education' ontology, and we compare the precision ratio which we get from four methods. Several experiment results refer to Table 2.

Figure 3 'Physical education' ontology O_3.

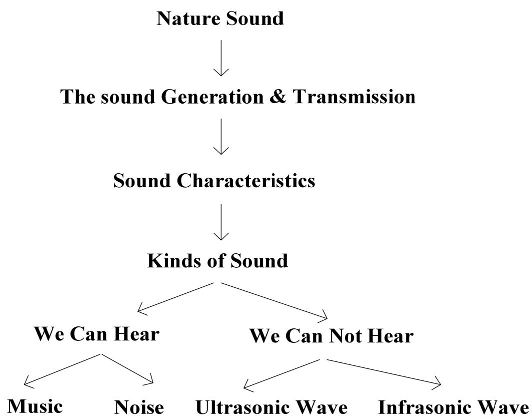


Table 2 The experiment data for ontology mapping

	P@1 average precision ratio	*P@3 average precision ratio*	*P@5 average precision ratio*
Our algorithm	70.97%	78.49%	92.90%
Algorithm in Huang et al. (2011a)	61.29%	73.12%	79.35%
Algorithm in Gao and Liang (2011)	69.13%	75.56%	84.52%
Algorithm in Gao et al. (2013)	67.74%	77.42%	89.68%

The experiment results in Table 2 reveal that our algorithm is more efficiently than algorithms raised in Huang et al. (2011a), Gao and Liang (2011) and Gao et al. (2013) especially when N is sufficiently large.

4.3 Experiment on plant data

In this subsection, 'PO' ontology O_4 which was constructed in http://www.plantontology.org (Figure 4 shows the basic structure of O_4) is used to test the efficiency of our new algorithm for ontology similarity measuring. The $P@N$ standard is used again for this experiment. We select 50 pairs of similarly vertices and 50 pairs of dissimilarly vertices, i.e., $|S| = |D| = 50$. Taking slack variable parameter $\gamma = 0.2$. Furthermore, we apply ontology method in Wang et al. (2010), Huang et al. (2011a) and Gao and Liang (2011) to the 'PO' ontology. Calculating the accuracy by these three algorithms and compare the result to algorithm rose in our paper, part of the data refer to Table 3.

Figure 4 'PO' ontology O_4

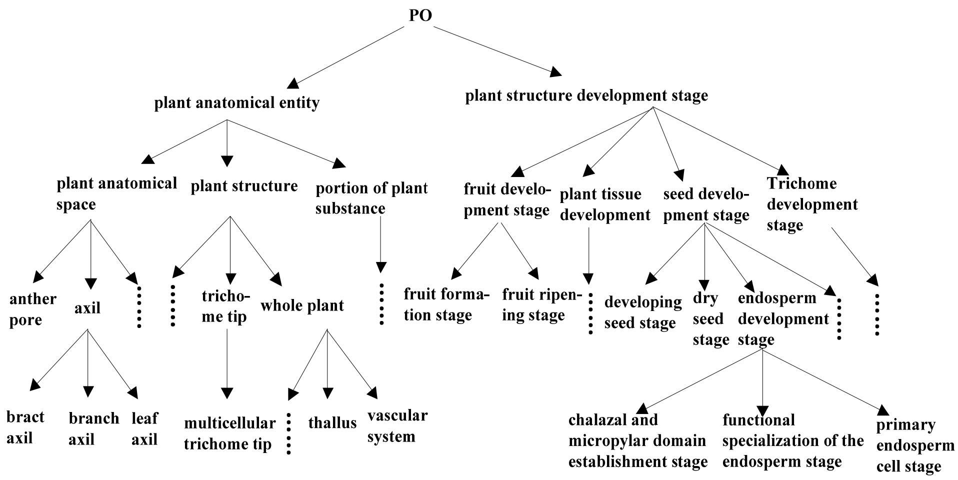


Table 3 The experiment data for ontology similairty measure

	P@3 average precision ratio	*P@5 average precision ratio*	*P@10 average precision ratio*
Our algorithm	48.63%	59.53%	74.19%
Algorithm in Wang et al. (2010)	45.49%	51.17%	58.59%
Algorithm in Huang et al. (2011a)	42.82%	48.49%	56.32%
Algorithm in Gao and Liang (2011)	48.31%	56.35%	68.71%

When N = 3, 5, or 10, the precision ratio in terms of our algorithm is higher than the precision ratio determined by algorithms proposed in Wang et al. (2010), Huang et al. (2011a) and Gao and Liang (2011). In particular, when N increases, such precision ratios are increasing apparently. Therefore, the algorithm described in our paper is superior to the method proposed by Wang et al. (2010), Huang et al. (2011a) and Gao and Liang (2011).

4.4 Experiment on humanoid robotics data

We use humanoid robotics ontologies O_5 and O_6 (constructed by Gao and Zhu (2014), and the structures of O_5 and O_6 are presented in Figures 5 and 6 respectively) for our last experiment. The goal of this experiment is to determine ontology mapping between O_5 and O_6 via similarity matrix which are deduced by Algorithm 1. *P@N* criterion is applied to measure the equality of the experiment. We only take five similarly pairs and five dissimilarly pairs in this experiment. Ontology algorithms in Gao and Lan (2011), Gao and Liang (2011) and Gao et al. (2013) are employed to humanoid robotics ontologies, and we compare the precision ratio which we get from four methods. Several experiment results refer to Table 4.

Figure 5 'Humanoid robotics' ontology O_5 (see online version for colours)

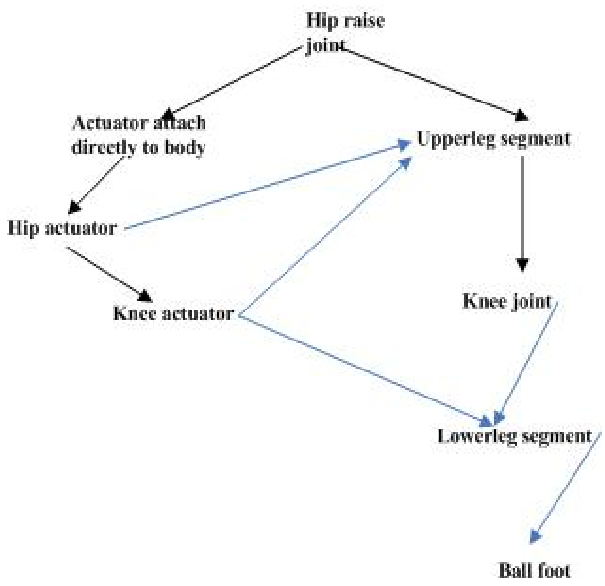


Figure 6 'Humanoid robotics' ontology O_6 (see online version for colours)

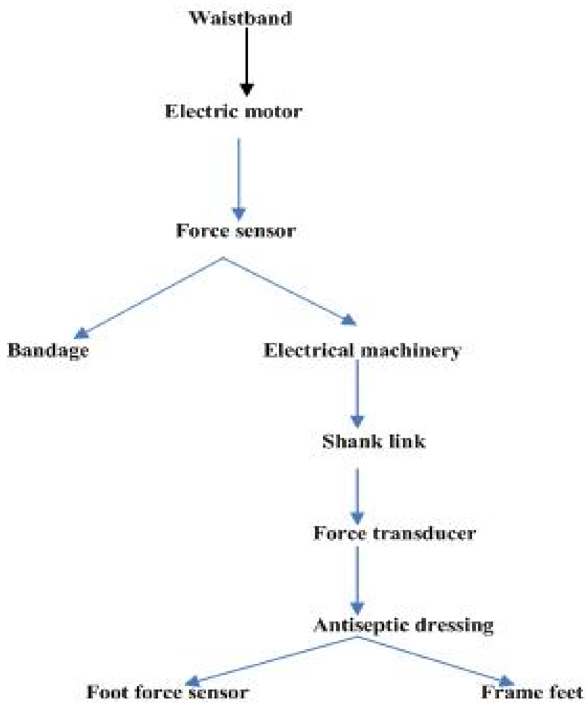


Table 4 The experiment data for ontology mapping

	$P@1$ average precision ratio	$P@3$ average precision ratio	$P@5$ average precision ratio
Our algorithm	27.78%	57.41%	74.44%
Algorithm in Gao and Lan (2011)	27.78%	48.15%	54.44%
Algorithm in Gao and Liang (2011)	22.22%	40.74%	48.89%
Algorithm in Gao et al. (2013)	27.78%	46.30%	53.33%

The experiment results in Table 4 reveal that our algorithm is more efficiently than algorithms raised in Gao and Lan (2011), Gao and Liang (2011) and Gao et al. (2013) especially when N is sufficiently large.

5 Conclusions

In this paper, we propose a new computation model for ontology similarity measure and ontology mapping application. The tricks are based on the metric distance learning and some new fashions are employed to get the optimal similarity matrix. At last, simulation data shows that our new algorithms have high efficiency in biology, physics education, plant science and humanoid robotics.

Acknowledgements

First we thank the reviewers for their constructive comments in improving the quality of this paper. This work was supported in part by science and technology of Jiangsu province (BE2011173). We also would like to thank the anonymous referees for providing us with constructive comments and suggestions.

References

Bouzeghoub, A. and Elbyed, A. (2006) 'Ontology mapping for web-based educational systems interoperability', *Interoperability in Business Information Systems*, Vol. 1, No. 1, pp.73–84.

Craswell, N. and Hawking, D. (2003) 'Overview of the TREC 2003 web track', *Proceeding of the Twelfth Text Retrieval Conference*, Gaithersburg, Maryland, NIST Special Publication, pp.78–92.

Davis, J. and Dhillon, S. (2006) 'Differential entropic clustering of multivariate Gaussians', *Adv. in Neural Inf. Proc. Sys. (NIPS)*, pp.337–344.

Fonseca, F., Egenhofer, E., Davis, C. and Camara, G. (2001) 'Semantic granularity in ontology-driven geographic information systems', *AMAI Annals of Mathematics and Artificial Intelligence – Special Issue on Spatial and Temporal Granularity*, Vol. 36, pp.121–151.

Gao, W. and Lan, M. (2011) 'Ontology mapping algorithm based on ranking learning method', *Microelectronics and Computer*, Vol. 28, No. 9, pp.59–61.

Gao, W. and Liang, L. (2011) 'Ontology similarity measure by optimizing NDCG measure and application in physics education', *Future Communication*, *Computing*, *Control and Management*, Vol. 142, No. 2, pp.415–421.

Gao, W. and Shi, L. (2013) 'Ontology similarity measure algorithm with operational cost and application in biology science', *BioTechnology: An Indian Journal*, Vol. 8, No. 11, pp.1572–1577.

Gao, W. and Xu, T. (2012) 'Characteristics of optimal function for ontology similarity measure via multi-dividing', *Journal of networks*, Vol. 7, No. 6, pp.1251–1259.

Gao, W. and Xu, T. (2013) 'Stability analysis of learning algorithms for ontology similarity computation', *Abstract and Applied Analysis*, 9pp [online] http://dx.doi.org/10.1155/2013/174802 (accessed 10 July 2013).

Gao, W. and Zhu, L. (2014) 'Gradient learning algorithms for ontology computing', *Computational Intelligence and Neuroscience*, 12pp [online] http://dx.doi.org/10.1155/2014/438291 (accessed 29 October 2014).

Gao, W., Gao, Y. and Zhang, Y. (2012) 'Strong and weak stability of *k*-partite ranking algorithm', *Information*, Vol. 15, No. 11A, pp.4585–4590.

Gao, W., Gao, Y. and Liang, L. (2013) 'Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping', *Journal of Chemical and Pharmaceutical Research*, Vol. 5, No. 9, pp.592–598.

Gao, W., Zhang, Y., Xu, T. and Zhou, J. (2012) 'Analysis for learning a similarity function with Ontology applications', *Journal of Information and Computational Science*, Vol. 9, No. 17, pp.5311–5318.

Gao, Y. and Gao, W. (2012) 'Ontology similarity measure and ontology mapping via learning optimization similarity function', *International Journal of Machine Learning and Computing*, Vol. 2, No. 2, pp.107–112.

Hu, B., Dasmahapatra, S., Lewis, P. and Shadbolt, N. (2003) 'Ontology-based medical image annotation with description logics', in *15th IEEE International Conference on Tools with Artificial Intelligence*, California, Sacramento, USA, pp.77–82.

Huang, X., Xu, T., Gao, W. and Gong, S. (2011b) '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, pp.571–574.

Huang, X., Xu, T., Gao, W. and Jia, Z. (2011a) 'Ontology similarity measure and ontology mapping via fast ranking method', *International Journal of Applied Physics and Mathematics*, Vol. 1, No. 1, pp.54–59.

Kulis, B., Sustik, M. and Dhillon, S. (2006) 'Learning low-rank kernel matrices', *International Conference on Machine Learning (ICML)*, Pittsburgh, PA, pp.505–512.

Lehmann, L. and Casella, G. (2003) *Theory of Point Estimation*, 2nd ed., Springer, Berlin.

Wang, Y., Gao W., Zhang, Y. and Gao, Y. (2010) 'Ontology similarity computation use ranking learning method', *The 3rd International Conference on Computational Intelligence and Industrial Application*, Wuhan, China, pp.20–22.

Yan, L., Gao, W. and Li, J. (2013) 'Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (II)', *Journal of Applied Science*, Vol. 13, No. 16, pp.3257–3262.

Zhu, L., Wu, F. and Gao, W. (2014) 'Theoretical analysis for new multi-dividing ontology algorithm based on AUC criterion', *Computing and Informatics*, In press.