http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (23): 7442-7447, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.7442.7447

Similarity Matrix Learning Using Dimensionality Reduction for Ontology Applications

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

Abstract: The aim of dimensionality reduction is to use low-dimensional data to represent high-dimensional data and thus can reduce the computational complexity. Graph spectral is an effective dimension reduction technology and used in various fields of computer science. In this study, we propose new algorithm for ontology similarity and ontology mapping application. The algorithm is given by calculating the similarity matrix in terms of spectral dimensionality reduction. Two experiments are designed to manifest the effectiveness of the algorithm.

Key words: Ontology, similarity measure, ontology mapping, spectral, dimensionality reduction, distance metric

INTRODUCTION

Ontology is defined as a shared conceptual and knowledge representation model which has been applied in collaboration, image retrieval, knowledge management, information systems, information retrieval search extension and intelligent information integration. Acted as an effective concept semantic model, ontology has been widely employed in many other disciplines such as social science (Bouzeghoub and Elbyed, 2006), biology medicine (Hu *et al.*, 2003) and geography science (Fonseca *et al.*, 2001).

Let G be a graph corresponding to ontology O. Each vertex on an ontology graph represents a concept; each edge on an ontology graph represents a connection between two concepts. The goal of ontology similarity measure is to approach a similarity function which maps each pair of vertices to a real number (or, to obtain a similarity matric). Let G_1, G_2,..., G_k be graphs corresponding to ontologies O_1, O_2,..., O_k, respectively. Set $G = G_1+G_2+...+G_k$. For every vertex $v \in V(G_i)$ ($1 \le i \le k$), the target of ontology mapping is searching similarity vertices from $G-G_i$. In this point of view, the essence of ontology mapping problem is just ontology similarity measure.

There are several effective technologies for ontology similarity measure and ontology mapping applications. Wang *et al.* (2010) proposed the ontology similarity calculation algorithm using ranking tricks. Huang *et al.* (2011a) raised fast ontology algorithm for reducing the complexity of the algorithm's implement. Gao and Liang (2012) argued that ontology function could be obtained by optimizing NDCG measure and applied such idea in physics education. Gao and Gao (2012) yielded the ontology function by virtue of the regression approach. Huang *et al.* (2011b) presented the ontology algorithm that ontology function is achieved based on half transductive ranking. Lan *et al.* (2012) explored the learning theory approach for ontology similarity computation in a setting where the ontology graph has tree structure. In terms of harmonic analysis and diffusion regularization on hypergraph, Gao *et al.* (2013a) presented new algorithms for ontology similarity measurement and ontology mapping. Very recently, Gao and Shi (2013) raised new ontology similarity algorithms such that the new computational models consider operational cost in the real implement.

Several studys contributed to the theoretical analysis for different ontology settings. Gao and Xu (2013) have studied the uniform stability of multi-dividing ontology algorithm and gave the generalization bounds for stable multi-dividing ontology algorithms. Gao *et al.* (2012) researched the strong and weak stability of multi-dividing ontology algorithm. Gao and Xu (2012) learned some characteristics for such ontology algorithm.

Gao *et al.* (2014) presented the characteristics of best ontology score function among piece constant ontology score functions. Gao *et al.* (2013b) and Yan *et al.* (2013) presented an approach of piecewise constant function approximation for AUC criterion multi-dividing ontology algorithms.

We are noticed that nonlinear dimensionality reduction and distance metric learning are two active tricks in ontology application. Xu *et al.* (2011) presented an ontology mapping algorithm based on dimensionality reduction theory. Lan *et al.* (2011) proposed a new ontology similarity method using graph spectral.

Corresponding Author: Yun Gao, Department of Editorial Yunnan Normal University, 650500, Kunming, Yunnan, China

However, these dimensionality reduction ontology algorithms were unsupervised, which implies they did not deal with the label information of the training data. In this study, we use the transductive framework of distance metric learning proposed by Li *et al.* (2007) for our ontology application. Such computation model directly connected with spectral dimensionality reduction methods and Laplacian eigenmaps. It only needs to solve a sparse eigenvalue problem. At last, the optimal similarity matrix D is learned and the similarity between vertices is obtained from D.

ONTOLOGY SIMILARITY MEASURE AND ONTOLOGY MAPPING ALGORITHMS

In this section, we first introduce some notations and problem setting and then the detailed technologies for ontology algorithms are presented. We use e, I and E to denote the vector of all ones, the identity matrix and the matrix of all ones, respectively. A^+ is denoted as the Moore-Penrose pseudo-inverse of a matrix A and diagonal matrix with its diagonal is x is denoted by diag(x).

For given n labeled vertices $\{(v_1, y_1),..., (v_n, y_n)\}$ and m unlabeled vertices $v_{n+1},..., v_{n+m}$ as sample set, the transductive distance metric learning problem as an optimization problem described as follows:

$$\min_D \frac{1}{n}\sum_{i=1}^{n} l(D, v_i, y_i) + \lambda\Omega(D) \qquad (1)$$

where, $l(D, v_i, y_i)$ is a loss function concerning D and labeled data (v_i, y_i), $\Omega(D)$ is a penalty function which used to control the smoothness of the optimal ontology matrix D and λ is a balance parameter controlling the strength of the empirical term and penalty term.

Note that the optimal distance metric D can be reduced to an $(n+m)\times(n+m)$ distance matrix with entries $d_{ij}=d^2(v_i, v_j)$. Let:

$$l(D, v_i, y_i) = \sum_{j=1}^{n}\sum_{k=1}^{n} l_{ijk} d_{jk},$$

$$\Omega(D) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} d_{ij}$$

and L_i and P be symmetric matrices whose coefficients are l_{ijk} and p_{ij} respectively. Hence, optimization problem (1) can be regard as matrix form:

$$\min_D Tr(\frac{1}{n}\sum_{i=1}^{n} L_i D + \lambda PD) \qquad (2)$$

An $n\times n$ matrix D is Euclidean if there exist n vertices v_i (i = 1,..., n) which can be embedded in a Euclidean space. Let d_{ij} be the squared Euclidean distance between vertices v_i and v_j. Since every Euclidean distance metric can be induced by an inner product. We have $d^2(v_i, v_j)= \langle v_i-v_j, v_i-v_j\rangle = \langle v_i- v_i \rangle+\langle v_j-v_j\rangle -2\langle v_i- v_i\rangle$ for certain inner product $\langle\cdot,\cdot\rangle$. Analogously, each Euclidean distance matrix can be induced by a Gram matrix Θ with its element $g_{ij} = \langle v_i-v_j\rangle$. It has determined by Gower and Legendre (1986) that: A matrix D is Euclidean if and only if it's associated Gram matrix:

$$\Theta = -\frac{1}{2}(I - \frac{1}{n}E)D(I - \frac{1}{n}E)$$

is positive semi-definite.

By virtue of above fact, (2) can be turned into an optimization problem concern the Gram matrix Θ. This implies, we need to search matrices L_j' and P' with characters $Tr(L_j\Theta) = Tr(L_iD)$ and $Tr(P'\Theta)= Tr(PD)$. The trick for constructing L_j' and P'' is heavily depended on the following proposition which proposed by Li *et al.* (2007):

- **Proposition 1:** (Li *et al.*, 2007) For symmetric matrix A, Euclidean matrix D and its corresponding Gram matrix Θ, we have $Tr(AD) = Tr(A'\Theta)$ if $A' = 2(diag(Ae)-A)$

After constructing L_i' and P', the computation framework (2) can be expressed into an optimization problem concern the Gram matrix Θ. Note that Θ has an orthogonal decomposition $\Theta = VV^T$ satisfies $Tr(\Theta) = Tr(V^TV) = 1$ since it is a positive semi-definite matrix. Thus, we obtain the following eigenvalue problem:

$$\begin{aligned} &\min_V Tr(V^T(\sum_{i=1}^{n} L_i' + \lambda P')V) \\ &s.t.\ V^TEV = 0 \\ &Tr(V^TV) = 1 \end{aligned} \qquad (3)$$

Then, the eigenvectors corresponding to the smallest eigenvalues of:

$$\sum_{i=1}^{n} L_i' + \lambda P'$$

is the optimal solution of V. As discussed in Li *et al.* (2007), the i-th row of V corresponds to the projection of v_i into a low-dimensional Euclidean space and the learned distance metric is the Euclidean metric in this transformed space. Hence, the technology is linked with spectral dimensionality reduction tricks.

The framework is summarized in the follows:

- Design the loss matrices L_i and penalty matrix P
- Construct L_j' and P' by virtue of Proposition 1
- Computer the eigenvectors of:

$$\sum_{i=1}^{n} L_i^{'} + \lambda P'$$

and take d eigenvectors corresponding to the smallest eigenvalues of the matrix, excluding the all-eigenvector corresponding to eigenvalue 0. Denote the matrix with each column an eigenvector by V

- The i-th row of V are the coordinates in a d-dimensional Euclidean space of the *i*-th item

A representer theorem: Let S: $H \rightarrow \mathbb{R}^{n+m}$ be the evaluation map on the labeled and unlabeled vertices, S (f) = (f (v_1), f (v_2), ..., f (v_{n+m})) and M be a positive semi-definite penalty matrix. Let H and $\tilde{H}$ be RKHSs associated with inner products $\langle \cdot,\cdot \rangle^{H}$ and $\langle \cdot,\cdot \rangle_{\tilde{H}} = \langle \cdot,\cdot \rangle_{H} + S(\cdot)^T MS(\cdot)$. Let k $\langle \cdot,\cdot \rangle$ be the representer of H and K be the (n+m)×(n+m) Gram matrix with $k_{ij} = k(v_i, v_j)$.

We use following regularization problem to optimize a loss function based on pairwise distances for one-dimensional embeddings:

$$\min_{f \in H} \sum_{i,j=1}^{n} w_{ij}(f(v_i) - f(v_j))^2 + \lambda \|f\|_{\tilde{H}}^2 \quad \text{s.t. } S(f)^T S(f) = 1 \tag{4}$$

where, w_{ij} are weights generated from label information controlling the loss on v_i and v_j. The following theorem is the main theoretical result of Li *et al.* (2007).

Theorem 1: (Li *et al.*, 2007) Assume that the Gram matrix K of rank n+m-1 satisfies Ke = 0. Let the penalty matrix and loss matrices in (3) satisfy:

$$P' = K + (I - \frac{E}{n})M(I - \frac{E}{n})$$

and:

$$\sum_{i=1}^{n} L_i^{'} = \sum_{i,j=1}^{n} w_{ij}(s_i - s_j)(s_i - s_j)^T$$

where s_i is a vector with 1 in the i-th position and 0 in other positions. If the one-dimensional optimal solution of (3) is v* and the optimal solution of (4) is f*, then we have v*= S(f*).

Design of the loss function: Intuitively, for distance-based technologies, distances between vertices with an edge associated should be smaller than distances between vertices with no edge associated. Consider labeled vertices set $\{(v_i, y_i)\}_{i=1,\ldots,n}$. Based on the intuition, the loss function L between two distances d_{ij} and d_{ik} is denoted as:

$$l(d_{ij}, d_{ik}) = \begin{cases} d_{ij} - d_{ik}, & y_i = y_j \text{ and } y_i \neq y_k \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Notice that the value of l (d_{ij}, d_{ik}) can be negative. We can resolve the problem in terms of specifying a maximum and minimum value of each d_{ij} and adding corresponding constants to the loss function.

To preserve intra-graph structures of ontology, the loss function only evaluated between neighboring vertices. For each (v_i, y_i), we take k intra neighbors and k inter-class neighbors of vertex in ontology graph. Let N_k^{in} (v_i) be the set of k intra neighbors of vertices v_i and N_k^{out} (v_i) be the set of k inter neighbors of v_i. We infer the loss function by averaging (5) over these neighbors as follows:

$$l(D, v_i, y_i) = \frac{1}{k} \sum_{v_j \in N_k^{in}(v_i)} d_{ij} - \frac{1}{k} \sum_{v_k \in N_k^{out}(v_i)} d_{ik}$$

Let the entries of symmetric matrix L be:

$$l_{ij} = \begin{cases} \frac{1}{k}, v_j \in N_k^{in}(v_i) \\ -\frac{1}{k}, v_j \in N_k^{out}(v_i) \end{cases}$$

Thus, we deduce that:

$$Tr(LD) = \sum_{i=1}^{n} L(D, v_i)$$

and the corresponding conditions in Theorem 1 also established for loss matrix.

If labeled vertices are very few in the training set, we may not be able to search k intra or inter neighbors for certain v_i and k. In this situation, (5) is averaged only on the available neighbor vertices and we should change the entries in the loss matrix L:

Design of the penalty matrix: In view of Theorem 1 which was determined by Li *et al.* (2007), we infer that the penalty matrix should adopt the form:

$$P = K^{+} + (I - \frac{1}{n}E)M(I - \frac{1}{n}E)$$

for certain positive semi-definite matrix M and Gram matrix K. We can choice K^+ = M, i.e., the Laplacian matrix of ontology graph G. However, generalizing new test vertices for the algorithm may become difficult. Alternatively, general kernel such as the Gaussian kernel could be used to compute K and its pseudo-inverse K^+. So, we use Laplacian matrix as penalty matrix.

EXPERIMENTS

In this section, we design two experiments concern ontology measure and ontology mapping.

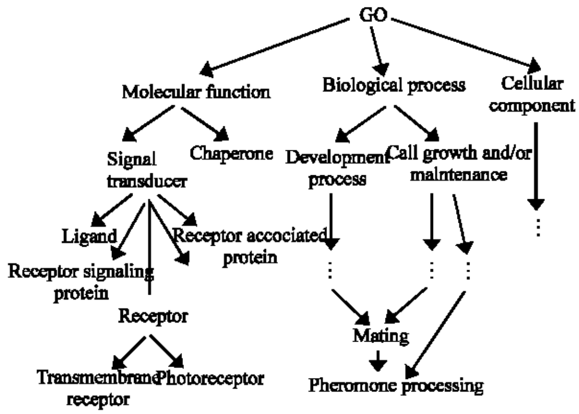

Fig. 1: "GO" Ontology O_1

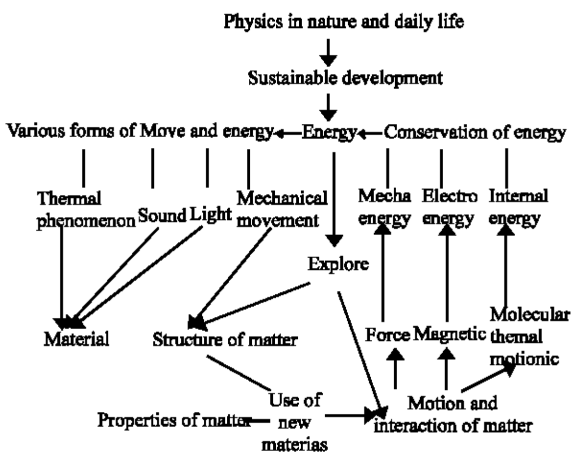

Fig. 2: "Physics Education" Ontology O_2

Table 1: 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 (%)
Algorithm in our study	56.71	64.44	75.63	85.37
Algorithm in Huang *et al.*, 2011a	47.73	55.62	69.93	76.82
Algorithm in Gao and Lan 2011	52.37	60.62	72.96	78.64

In the first experiment, we use biology ontology O_1 which was constructed in http: //www. geneontology. org. Fig. 1 shows O_1. P@ N (Precision Ratio, see Craswell and Hawking 2003 for more detail) is employed to quantify the quality of our experiment. We first give 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 Huang *et al.* (2011a) and Gao and Lan (2011) to the "GO" ontology. We calculate the accuracy by these two algorithms and compare the result to algorithm proposed in our study, part of the data refer to Table 1.

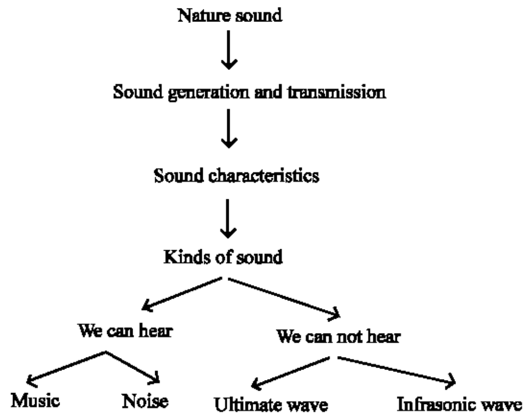

Fig. 3: "Physics Education" Ontology O_3

Table 2: Experiment results of ontology mapping

	P@1 average precision ratio (%)	P@3 average precision ratio (%)	P@5average precision ratio (%)
Algorithm in our study	48.39	62.37	71.61
Algorithm in Huang *et al.* (2011a)	41.94	49.46	59.35
Algorithm in Gao and Lan (2011)	45.16	56.99	64.52

For the second experiment, we use "Physics Education" ontology O_2 and O_3, as Fig. 2 shows O_2 and Fig. 3 shows O_3. The aim of this experiment is to obtain the ontology mapping between O_2 and O_3. Again, P@N is used to measure the quality of experiment. Also, we apply ontology algorithms in Huang *et al.* (2011b) and Gao and Lan (2011) on "Physics Education" ontology and compare the precision ratio which yield from these methods. Some results refer to Table 2.

From the experiment results display above, we arrived at the conclusion that our algorithm is more efficiently than algorithms raised in Huang *et al.* (2011b) and Gao and Lan (2011) especially when N is lager enough. Therefore, this new ontology similarity algorithm by virtue of distance metric learning and spectral dimensionality reduction has high efficiency.

The experiment data in Table 2 showed that our algorithm is more efficiently than algorithms raised in Huang *et al.* (2011b) and Gao and Lan (2011) especially when N is lager. Hence, new ontology mapping algorithm based on distance metric learning by spectral dimensionality reduction is effective in certain special applications.

CONCLUSION

Ontology, as a data representation model, has been widely used in various fields and proved to have a high efficiency. In our study, we apply the tricks of distance metric learning and spectral dimensionality reduction to design the new ontology algorithms. The new algorithms have high quality according to the simulation data presented in above section. This contributes to the state of art for ontology applications.

ACKNOWLEDGMENTS

First we thank the reviewers for their constructive comments in improving the quality of this study. 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, Key Science and Technology Research Project of Education Ministry 210210 and PHD initial funding of the third author in this study. We also would like to thank the anonymous referees for providing us with constructive comments and suggestions.

REFERENCES

Bouzeghoub, A. and A. Elbyed, 2006. Ontology mapping for web-based educational systems interoperability. Interoperability Bus. Inform. Syst., 1: 73-84.

Craswell, N. and D. Hawking, 2003. Overview of the TREC 2003 web track. Proceedings of the 12th Text Retrieval Conference, November 18-21, 2003, Gaithersburg, MD., USA., pp: 78-92.

Fonseca, F., E. Egenhofer, C. Davis and G. Camara, 2001. Semantic granularity in ontology-driven geographic information systems. Ann. Math. Artif. Intell., 36: 121-151.

Gao, W. and L. Liang, 2012. Ontology Similarity Measure by Optimizing NDCG Measure and Application in Physics Education. In: Future Communication, Computing, Control and Management, Zhang, Y. (Ed.). Vol. 2, Springer, New York, USA., ISBN-13: 9783642273131, pp: 415-421.

Gao, W. and L. Shi, 2013. Ontology similarity measure algorithm with operational cost and application in biology science. BioTechnol.: Indian J., (In Press).

Gao, W. and M.H. Lan, 2011. Ontology mapping algorithm based on ranking learning method. Microelectron. Comput., 9: 59-61.

Gao, W. and T. Xu, 2012. Characteristics of optimal function for ontology similarity measure via multi-dividing. J. Networks, 7: 1251-1259.

Gao, W. and T. Xu, 2013. Stability analysis of learning algorithms for ontology similarity computation. Abstr. Applied Anal. 10.1155/2013/174802.

Gao, W., L. Yan and L. Liang, 2013. Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (I). Int. J. Comput. Appl. Technol., (In Press).

Gao, W., T. Xu, J. Gan and J. Zhou, 2014. Linear statistical analysis of multi-dividing ontology algorithm. J. Inform. Comput. Sci., (In Press).

Gao, W., Y. Gao and L. Liang, 2013. Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping. J. Chem. Pharm. Res., 5: 592-598.

Gao, W., Y. Gao and Y. Zhang, 2012. Strong and weak stability of *k*-partite ranking algorithm. Information, 15: 4585-4590.

Gao, Y. and W. Gao, 2012. Ontology similarity measure and ontology mapping via learning optimization similarity function. Int. J. Mach. Learn. Comput., 2: 107-112.

Gower, J.C. and P. Legendre, 1986. Metric and Euclidean properties of dissimilarity coefficients. J. Classification, 3: 5-48.

Hu, B., S. Dasmahapatra, P. Lewis and N. Shadbolt, 2003. Ontology-based medical image annotation with description logics. Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, November 3-5, 2003, California, Sacramento, USA., pp: 77-82.

Huang, X., T. Xu, W. Gao and S. Go, 2011. Ontology similarity measure and ontology mapping using half transductive ranking. Proceedings of the 4th IEEE International Conference on Computer Science and Information Technology, June 10-12, 2011, Chengdu, China, pp: 571-574.

Huang, X., T. Xu, W. Gao and Z. Jia, 2011. Ontology similarity measure and ontology mapping via fast ranking method. Int. J. Applied Phys. Math., 1: 54-59.

Lan, M.H., J. Xu and Y. Sun, 2011. Ontology similarity computation based on spectral graph theory. Comput. Eng. Appl., 47: 110-112.

Lan, M.H., Y.J. Ren, J. Xu and W. Gao, 2012. Ontology similarity computation using *k*-partite ranking method. J. Comput. Appl., 32: 1094-1096.

Li, F., J. Yang and J. Wang, 2007. A transductive framework of distance metric learning by spectral dimensionality reduction. Proceedings of the 24th International Conference on Machine Learning, June 20-24, 2007, Corvallis, OR., USA., pp: 513-520.

Wang, Y., W. Gao, Y. Zhang and Y. Gao, 2010. Ontology similarity computation use ranking learning method. Proceedings of the 3rd International Conference on Computational Intelligence and Industrial Application, November 6-7, 2010, Wuhan, China, pp: 20-22.

Xu, T.W., J.H. Gan, L.J. Gao and W. Gao, 2011. Ontology mapping algorithm based on dimensionality reduction theory. J. Northwest Normal Univ. (Nat. Sci.), 47: 52-55.

Yan, L., W. Gao and J.S. Li, 2013. Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (II). J. Applied Sci., 13: 3257-3262.