Bilinear model for ontology mapping

Jian-Zhang Wu[1, 2*], Yu Xiao[3], Wei Gao[4]

[1]*School of Computer Science & Engineer, Southeast University, Nanjing 210096, China*

[2]*Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing 210096, China*

[3]*School of Continuing Education, Southeast University, Nanjing 210096, China*

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

Received 15 July 2014, www.tsi.lv

Abstract

As a model of concept representation, ontology has widely applied to various disciplines. Ontology mapping is used to create the link between different ontologies. In this paper, we present a new ontology mapping algorithm by virtue of bilinear model. The linear mapping pair is given by the iterative procedure. Two strategies are manifested to obtain the finally ontology mapping. The simulation experimental results show that the proposed new technologies have high accuracy and efficiency on ontology mapping in certain applications.

Keywords: ontology, ontology mapping, linear mapping, bilinear model, dimensionality reduction

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 is also employed in disciplines beyond computer science, such as social science (for instance, see [1]), biology science [2] and geography science [3].

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 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 to 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 corresponding to vertices in S_v are semantically close to the concept corresponding to v. One method to get such mapping is, for each $v\in G_1$, to compute the similarity $S(v,v_j)$ where $v_j \in V(G_2)$ and to 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 an optimal similarity function S from dimensionality reduction standpoint. In fact, our approach for obtaining such similarity function is based on the linear mapping pair.

For ontology similarity measure, there are several effective learning tricks. Wang et al. [4] 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., [5] presented a fast ontology algorithm for calculating the ontology similarity in a short time. Gao and Liang [6] raised that the optimal ontology function can be determined by optimizing NDCG measure, and applied such idea in physics education. Gao and Gao [7] deduced the ontology function using the regression approach. Huang et al., [8] obtained ontology similarity function based on half transductive learning. Gao and Xu [9] explored the learning theory approach for ontology similarity computation using k-partite ranking method. Zhu and Gao [10] proposed a new criterion for ontology computation from AUC and multi-dividing standpoint. Gao et al., [11] presented a new ontology mapping algorithm using harmonic analysis and diffusion regularization on hypergraph. Very recently, Gao and Shi [12] proposed a new ontology similarity computation technology such that the new calculation model considers operational cost in the real implement.

In this paper, we determine the new ontology mapping algorithm based on dimensionality reduction idea and bilinear learning model. Using the optimization algorithm, we determine the linear mapping (L_1, L_2) to compute the similarity of vertices from two ontologies. The experiment is designed to show the efficiency of the algorithm.

2 Model and algorithm

For each vertex v, we use a vector to represent all its information. For two ontologies O_1 and O_2, their structures can be determined by two ontology graphs

* ***Corresponding author*** e-mail: jzwu@njnet.edu.cn

$G_1=(V_1,E_1)$ and $G_2=(V_2,E_2)$ respectively. Suppose that $V_1 \subset \mathbb{R}^{d_1}$ and $V_2 \subset \mathbb{R}^{d_2}$. That is to say, we use a vector with dimension d_1 to represent the information of vertex in V_1 and use a vector with dimension d_2 to represent the information of vertex in V_2. For any $v_i \in V_1$, and $v_j \in V_2$, $S(v_i,v_j)=S_{ij}$ indicates the similarity between concepts corresponding to v_i and v_j. Our goal is to learn an optimal similarity function S based on the sample triple $D=\{(v_i,v_j,S_{ij})\}$, where $v_i \in V_1$, $v_j \in V_2$. For such triple D, let $V_1^D=\{v_i\}\subseteq V_1$, $V_2^D=\{v_j\}\subseteq V_2$, $n_1=|V_1^D|$ and $n_2=|V_2^D|$.

We are interested in searching a linear mapping pair (L_1,L_2) such that the corresponding images $L_1^T v_i$ and $L_2^T v_j$ are in the same d-dimensional latent space L with $d \ll \min\{d_1,d_2\}$ and the degree of similarity between ontology vertices $v_i \in V_1$ and $v_j \in V_2$ can be reduced to L's dot product:

$$D_{L_1,L_2}(v_i,v_j)=v_i^T L_1 L_2^T v_j .$$

By virtue of the trick used in [13] for kernel learning, we aim to maximize the following expected version:

$$E_{v_i,v_j}\{S(v_i,v_j)D_{L_1,L_2}(v_i,v_j)\}= E_{v_i}E_{v_j|v_i}\{S(v_i,v_j)v_i^T L_1 L_2^T v_j\}. \quad (1)$$

The Equation (1) could be estimated as follows:

$$\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} S_{ij} v_i^T L_1 L_2^T v_j .$$

Hence, the ontology mapping problem is boiled down to

$$\arg\max_{L_1,L_2}\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} S_{ij} v_i^T L_1 L_2^T v_j , \quad (2)$$

s.t. $L_1 \in H_1$, $L_2 \in H_2$, where H_1 and H_2 are the hypothesis spaces for L_1 and L_2 respectively. Since the final computational model is linear in view of both ontology vertices v_i and v_j, learning model (2) is actually a bilinear model for calculating similarity in two spaces.

We apply l_1 norm and l_2 norm constraints on L_1 and L_2. Let $|\cdot|$ and $\|\cdot\|$ be l_1-norm and l_2-norm respectively, and $l_{v_i x}$ and $l_{v_j y}$ be the x-th and y-th row of L_1 and L_2. Specifically, we introduce two hypothesis spaces as:

$$H_1=\{L_1 \mid |l_{v_i x}|\le\lambda_{v_i}, \|l_{v_i x}\|\le\theta_{v_i}, x=1,\dots,d_1\},$$

$$H_2=\{L_2 \mid |l_{v_j y}|\le\lambda_{v_j}, \|l_{v_j y}\|\le\theta_{v_j}, y=1,\dots,d_1\},$$

where $\{\lambda_{v_i},\theta_{v_i},\lambda_{v_j},\theta_{v_j}\}$ are parameters selected by experts. Here the constraints relying on l_1-norm will induce row-wise sparsity in L_1 and L_2. Furthermore, the l_2-norm on rows with regularization can avoid degenerative solutions. By virtue of the definition of H_1 and H_2, we infer the following program:

$$\arg\max_{L_1,L_2}\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} S_{ij} v_i^T L_1 L_2^T v_j , \quad (3)$$

s.t.:

$$|l_{v_i x}|\le\lambda_{v_i}, \|l_{v_i x}\|\le\theta_{v_i}, |l_{v_j y}|\le\lambda_{v_j}, \|l_{v_j y}\|\le\theta_{v_j}, \quad x\in\{1,\dots,d_1\}, y\in 1,\dots,d_2 .$$

In practice reality, we solve the following variant version of Equation (3) for easier computation

$$\arg\min_{L_1,L_2} -\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} S_{ij} v_i^T L_1 L_2^T v_j + \beta\sum_{x=1}^{d_1}|l_{v_i x}|+\gamma\sum_{y=1}^{d_2}|l_{v_j y}|, \quad (4)$$

s.t. $\|l_{v_i x}\|\le\theta_{v_i}$, $\|l_{v_j y}\|\le\theta_{v_j}$, $x\in\{1,\dots,d_1\}$, $y\in 1,\dots,d_2$,

where $\beta > 0$ and $\gamma > 0$ are the balance parameters to control the trade-off between objective term and penalty term. For given L_2, the objective mapping of Equation (4) can be re-represented as:

$$\sum_{x=1}^{d_1}\left\{\left(-\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} v_i^x S_{ij} L_2^T v_j\right)^T l_{v_i x}+\beta|l_{v_i x}|\right\}.$$

By using $\boldsymbol{\omega}_x=[\omega_x^1,\omega_x^2,\dots,\omega_x^d]^T$ to represent the d-dimensional $\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} v_i^x S_{ij} L_2^T v_j$, we infer the optimal $l_{v_i x}$ as:

$$(l_{v_i x}^k)^*=C_{v_i}\left(\max\left(|\omega_{v_i}^k|-\beta,0\right)\operatorname{sign}(\omega_{v_i}^k)\right),\ k\in\{1,\dots,d\}, \quad (5)$$

where $l_{v_i x}^k$ is the k-th element of $l_{v_i x}$:

$$\operatorname{sign}(x)=\begin{cases}1, & x>0\\ 0, & x=0\\ -1, & x<0\end{cases},$$

and C_{v_i} is a constant which makes $\|l_{v_i x}^*\|=\theta_{v_i}$ if there exist non-zero elements in $l_{v_i x}^*$, and $C_{v_i}=0$ otherwise.

For given L_1, the objective mapping of Equation (4) can similarly re-written as:

$$\sum_{y=1}^{d_2}\left\{\left(-\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} v_j^y S_{ij} L_1^T v_i\right)^T l_{v_j y} + \gamma\left|l_{v_j y}\right|\right\}.$$

In terms of the same fashion, we use $\mathbf{\eta}_y = \left[\eta_y^1, \eta_y^2, ..., \eta_y^d\right]^T$ to represent the d-dimensional $\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} v_j^y S_{ij} L_1^T v_i$, we yield the optimal $l_{v_j y}$ as:

$$\left(l_{v_j y}^k\right)^* = C_{v_j}(\max(\left|\eta_{v_j}^k\right| - \gamma, 0)\operatorname{sign}(\eta_{v_j}^k)),\ k\in\{1,...,d\}, \quad (6)$$

where $l_{v_j y}^k$ is the k-*th* element of $l_{v_j y}$, and C_{v_j} is a constant which makes $\left\|l_{v_j y}^*\right\| = \theta_{v_j}$ if there exist non-zero elements in $l_{v_j y}^*$, and $C_{v_j} = 0$ otherwise.

Let:

$$w_{v_i x} = \frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} v_i^x S_{ij} v_j$$

and

$$w_{v_j y} = \frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} v_j^y S_{ij} v_i ,$$

which does not depend on the change of L_1 and L_2, and can be pre-calculated. It is easy to verify that: $\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} v_i^x S_{ij} L_2^T v_j = L_2^T w_{v_i x} =$

and

$$\frac{1}{n_1 n_2}\sum_{v_i\in V_1^D}\sum_{v_j\in V_2^D} v_j^y S_{ij} L_1^T v_i = L_1^T w_{v_j y} .$$

Let N_{v_i} be the average number of non-zeros in all v_i per dimension and N_{v_j} be the average number of non-zeros in all v_j per dimension, $\tilde{n}_1$ be the average number of related v_i samples per v_j and $\tilde{n}_2$ be the average number of related v_j samples per v_i, c_1 be the average number of non-zeros in each v_i sample and c_2 be the average number of non-zeros in each v_j sample. Now, we present the following two algorithms:

Algorithm 1. Calculating $w_{v_i x}$ and $w_{v_j y}$

Input $D=\{(v_i,v_j,S_{ij})\}$, where $1\le i\le n_1$ and $1\le j\le n_2$.

For x=1:d_1, $w_{v_i x}\leftarrow 0$; For y=1:d_2, $w_{v_j y}\leftarrow 0$.

For x=1:d_1, i=1:n_1, j=1:d_2, $w_{v_i x}\leftarrow w_{v_i x} + \frac{1}{n_1 n_2} v_i^x S_{ij} v_j$.

For y=1:d_2, i=1:n_1, j=1:d_2, $w_{v_j y}\leftarrow w_{v_j y} + \frac{1}{n_1 n_2} v_j^y S_{ij} v_i$.

Output: $\{w_{v_i x}\}_{x=1}^{d_1}$ and $\{w_{v_j y}\}_{y=1}^{d_2}$.

Algorithm 2. Calculating L_1^t and L_2^t.

Input $\{w_{v_i x}\}_{x=1}^{d_1}$, $\{w_{v_j y}\}_{y=1}^{d_2}$, d, β, γ, θ_1, θ_2, and set L_1^0 and L_2^0 randomly $t\leftarrow 0$.

While $t\le T$. For x=1:d_1, compute $\mathbf{\omega}_x$ using $(L_1^t)^T w_{v_i x}$ and determine $(l_{v_i x})^*$ in terms of Equation (5), update L_1^{t+1};

For y=1:d_2, compute $\mathbf{\eta}_y$ using $(L_2^t)^T w_{v_j y}$ and determine $(l_{v_j y})^*$ in terms of Equation (6), update L_2^{t+1}; $t\leftarrow t+1$.

Output L_1^t and L_2^t.

The complexities of Algorithm 1 and Algorithm 2 are $O(d_1 N_{v_i}\tilde{n}_2 c_2 + d_2 N_{v_j}\tilde{n}_1 c_1)$ and $O(d_1 W_1 d + d_2 W_2 d)$ respectively, where W_1 is the number of non-zeros for each $w_{v_i x}$ on average and W_2 is the number of non-zeros for each $w_{v_j y}$ on average.

After the similarity between vertices are determined by bilinear model, we select a strategy to derive finally ontology mapping. Following two strategies could be used for getting ontology mapping.

Strategy 1. For each $v\in V(G_i)$, i=1,2. Let $N\in\mathbb{N}$ be a parameter, and:

$$v_1 = \max_{v'\notin V(G_i)}\left\{S(v,v')\right\},$$

$$v_2 = \max_{v'\notin V(G_i), v'\ne v_1}\left\{S(v,v')\right\},$$

$$v_3 = \max_{v'\notin V(G_i), v'\ne v_1, v'\ne v_2}\left\{S(v,v')\right\},$$

$$v_N = \max_{v'\notin V(G_i), v'\ne v_1, v'\ne v_2,\cdots,v'\ne v_{N-1}}\left\{S(v,v')\right\}.$$

Then, we deduce:

$$map(v) = \{v_1, v_2, \cdots, v_N\}.$$

Strategy 2. For each $v\in V(G_i)$, i=1,2. Let $M\in\mathbb{R}^+$ be a parameter, and

$$map(v) = S\left\{(v,v')\ge M\left|v'\notin V(G_i)\right.\right\}.$$

3 Experiment

Experiment of relevance ontology mapping is designed below. In order to adjacent to the setting of ontology algorithm, we use a vector 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.

We use physical education ontologies O_1 and O_2 (the structures of O_1 and O_2 are presented in Figures 1 and 2 respectively) for our experiment. The goal of this experiment is to determine ontology mapping between O_1

and O_2 via linear mapping L_1 and L_2 which is deduced by Algorithm 1 and Algorithm 2. *P@N* criterion (Precision Ratio, see Craswell and Hawking, [14]) 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 [11, 5] and [6] are employed to "physical education" ontology, and we compare the precision ratio which we get from four methods. Several experiment results refer to Table 1.

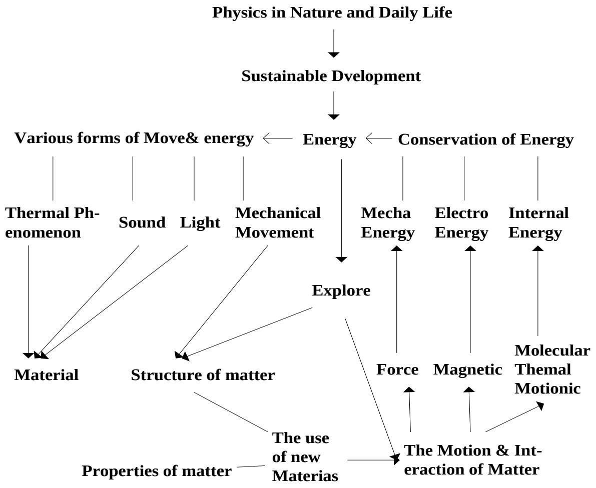

FIGURE 1 "Physical Education" Ontology O_1

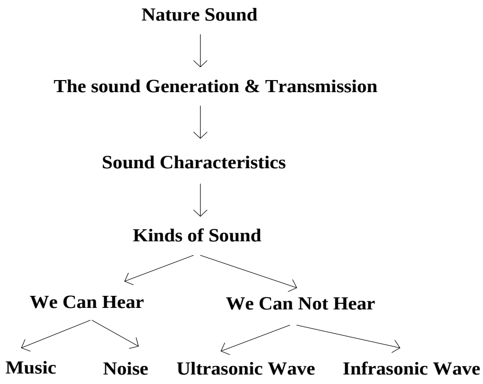

FIGURE 2 "Physical Education" Ontology O_2

The experiment results in Table 1 reveal that the precision ratio in terms of our algorithm higher than the precision ratio determined by algorithms proposed in [11, 5] and [6] by taking *N*= 1, 3 or 5. Specially, as *N* becomes large, such precision ratios in terms of our algorithm are increasing apparently. In this point of view, our algorithm is more efficient than algorithms raised in [11, 5] and [6] especially when *N* is sufficiently large.

TABLE 1 The experiment data of ontology mapping

	***P*@1 average precision ratio**	***P*@3 average precision ratio**	***P*@5 average precision ratio**
Algorithm presented in our paper	70.97%	79.37%	90.48%
Algorithm presented in [11]	67.74%	77.42%	89.68%
Algorithm presented in [5]	61.29%	73.12%	79.35%
Algorithm presented in [6]	69.13%	75.56%	84.52%

4 Conclusions

In this paper, we propose a new computation model for ontology mapping application. The model is bilinear and the algorithm is essentially a kind of dimensionality reduction algorithm which maps the high-dimensional ontology space into low-dimensional. At last, simulation data shows that our new algorithm has high efficiency in physics education ontologies. The algorithm achieved in our paper illustrates the promising application prospects for ontology mapping. The technologies raised in our paper contribute to the state of the art.

Acknowledgment

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), and Key Laboratory of Computer Network and Information Integration founding in Southeast University. We also would like to thank the anonymous referees for providing us with constructive comments and suggestions.

References

[1] Bouzeghoub A, Elbyed A 2006 Ontology mapping for web-based educational systems interoperability *Journal of Interoperability in Business Information Systems* **1**(1) 73-84

[2] Hu B, Dasmahapatra S, Lewis P, Shadbolt N 2003 Ontology-based medical image annotation with description logics *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence* California Sacramento USA 77-82

[3] Fonseca F, Egenhofer E, Davis C, 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* 36 121-51

[4] Wang Y, Gao W, Zhang Y, Gao Y 2010 Ontology similarity computation use ranking learning Method *The 3rd International Conference on Computational Intelligence and Industrial Application* Wuhan China 20-2

[5] Huang X, Xu T, Gao W, Jia Z 2011 Ontology similarity measure and ontology mapping via fast ranking method *International Journal of Applied Physics and Mathematics* **1**(1) 54-9

[6] Gao W, Liang L 2011 Ontology similarity measure by optimizing NDCG measure and application in physics education, *Future Communication, Computing, Control and Management* 142 415-21

[7] Gao Y, Gao W 2012 Ontology similarity measure and ontology mapping via learning optimization similarity function *International Journal of Machine Learning and Computing* **2**(2) 107-12

[8] Huang X, Xu T, Gao W, Gong S 2011 Ontology similarity measure and ontology mapping using half transductive ranking *Proceedings of 4th IEEE international Conference on Computer Science and Information technology* Chengdu China 571-4

[9] Gao W, Xu T 2013 Stability analysis of learning algorithms for ontology similarity computation Abstract and Applied Analysis 2013 Article ID 174802

[10] Zhu L, Gao W Theoretical Analysis for New Multi-dividing Ontology Algorithm Based on AUC Criterion, Manuscript *(in Chinese)*

[11] Gao W, Gao Y, Liang L 2013 Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping *Journal of Chemical and Pharmaceutical Research* **5**(9) 592-8

[12] Gao W, Shi L 2013 Ontology similarity measure algorithm with operational cost and application in biology science *An Indian Journal of BioTechnology (in Chinese)*

[13] Cristianini N, Shawe-taylor J, Elisseeff A, Kandola J 2011 On kernel-target alignment *Proceedings of the Neural Information Processing Systems NIPS'01* MIT Press 367-73

[14] Craswell N, Hawking D, Wilkinsin R, Wu M 2003 Overview of the TREC 2003 web track *Voorheer E (ed.) NIST Special Publication 500-255: The Twelfth Text REtrieval Conference (TREC 2003)* Gaithersburg MD 78-92

Authors

Jian-Zhang Wu

Current position, grades: Associate Professor at the School of Computer Science & Engineer (Southeast University).
Scientific interests: cloud computing, information retrieval, machine learning.

Yu Xiao

Current position, grades: Lecturer at the School of Continuing Education (Southeast University).
Scientific interests: web technology, cloud computing, educational technology.

Wei Gao

Current position, grades: Lecturer at School of School of Information Science and Technology (Yunnan Normal University).
Scientific interests: information retrieval, graph theory, machine learning.