Ontology Similarity Computation and Ontology Mapping Using Distance Matrix Learning Approach

Meihui Lan, Jian Xu, and Wei Gao

***Abstract*—Ontology, a common tool in various fields of natural sciences, aims to get the optimal ontology similarity calculation function. Favored by researchers from semantic query and other disciplines, it is used to calculate the similarity between ontology concepts. The semantic information of each concept is expressed by a *d*-dimensional vector, and the similarity calculation is transformed into the geometric distance calculation of the two corresponding vectors. Using the Mahalanobis distance calculation formula, the ontology algorithm can contribute to getting the optimal distance matrix. In this paper, in terms of the coordinate descent trick and iterative method, we get the ontology distance matrix learning algorithm, and then apply it to ontology similarity computation and ontology mapping. Moreover, the ontology distance matrix learning approach in the manifold setting is discussed, and its kernel solution is studied as well. The main ontology learning algorithm is illustrated by a comparison of other ontology algorithmic data in a specific application context.**

***Index Terms*—Ontology, similarity measure, ontology mapping, distance matrix, iterative**

I. Introduction

THE term ontology, derived from philosophy, usually describes the essential association between things. Serving as a data structure representation model in computer field, it has been widely used in various fields of computer science. Ontology similarity calculation and ontology mapping algorithm has become one of the core contents of the research of knowledge representation. At the same time as a conceptual structure, ontology has been widely used in biomedical (Mork and Bernstein [1]), geography (Fonseca et al. [2]), physics and mathematics (Elizarov et al. [3]), social science (Ochara [4]) and many other fields. For ontology mapping, its essence is to calculate the similarity between concepts from different ontologies. Therefore, the ontology similarity computation algorithm can be an ontology mapping algorithm after a proper conversion.

In the information retrieval, the vertices in ontology represent concepts, and the edges denote the correlation between concepts. The set of concepts corresponding to all vertices B's satisfying $Sim(A,B)>M$ is returned to the user as a query extension of the corresponding concept of vertex A. For the ontology mapping, set graphs $G_1,\cdots,G_m$ correspond to the main ontologies $O_1,\cdots,O_m$ respectively. For each $A\in V(G_i)$, where $1\le i\le m$, in the $G-G_i$, find all the set of vertex *B* corresponding concepts satisfying $Sim(A,B)>M$ returns to the user as the query extension of the corresponding concept of vertex *A*.

The ontology is applied as a tool to various fields. For example, in the biological field, the GO ontology (http: // www. geneontology.org) contains information on cellular, molecular and biological processes, with about 23,700 terms and more than 16 million in 20 biological databases made comment on the gene. The analysis of the ontology can help biologists understand the interrelated characteristics of genes between different biological databases.

In recent years, the ontology has been applied in various fields. Przydzial [5] applied the bulk to protein retrieval in pharmaceuticals. Koehler et al. [6] used the ontology to the characterization database between molecules and disease. Ivanovic and Budimac [7] reviewed the application of ontology in the medical field. Hristoskova [8] made use of the ontology to the creation of a personal care system. Kabir et al [9] established an effective social information management platform by means of the ontology data structure. Ma et al. [10] proposed an ontology model framework based on stable semantic retrieval. Li et al. [11] obtained a new ontology data representation model and applied it to customer shopping systems. Santodomingo et al. [12] proposed a matching system for ontology domain expert knowledge. Pizzuti et al. [13] innovated the food ontology and gave some practical application of the ontology. Lasierra et al. [14] suggested that the ontology can be applied to the design of the building and applied to the design and maintenance of the patient's home. Carlini and Makowski [15] applied the gene GO ontology to the study of the preferred codeword in insect homology. Nicolai [16] described the theory of deflation and its ontology representation. Correae et al. [17] incorporated the ontology approach into a modular science-based annotation science document system and applied it to the field of drug and infectious disease control. Duran-Limon et al. [18] proposed a method of deriving product architecture based on ontology.

Manuscript received June 10, 2017; revised October 30, 2017. This work was supported in part by the Key Laboratory of Educational Informatization for Nationalities, Ministry of Education, the National Natural Science Foundation of China (60903131).

M. H. Lan is with Department of Computer Science and Engineering, Qujing Normal University, Qujing, 655011, China, e-mail: lanhuimei97@126.com.

J. Xu is with Department of Computer Science and Engineering, Qujing Normal University, Qujing, 655011, China, e-mail: xujian@163.com.

W. Gao is with the School of Information and Technology, Yunnan Normal University, Kunming, 650500, China, e-mail: gaowei@ynnu.edu.cn.

Chabot et al. [19] obtained a time scale reconstruction and analysis method for digital events based on ontology. Elbers and Taylor [20] gave the ontology-based gene workflow algorithm and applied it to the targeted region of the targeted enrichment experiment. Rani et al. [21] obtained an ontology-based adaptive, personalized learning system that was implemented simultaneously with cloud storage technology through software proxies. Sangiacomo [22] studied the role of the ontology in the behavioral determinism system and obtained several results. Azevedo et al. [23] built a model of the analysis of the resources and capabilities of the enterprise architecture modeling under the ontology framework. Wimmer and Rada [24] established an ontology-based analysis and evaluation algorithm to determine the quality of the information. Trokanas and Cecelja [25] discussed an ontology evaluation system in the field of chemical process systems engineering. Chhun et al. [26] introduced QoS ontology for service selection and reuse, which focused on assessing the quality of the service. Costa et al. [27] constructed the corresponding ontology algorithm for knowledge sharing and reused in the construction sector. Panov et al. [28] studied the gene ontology OntoDT for the representation of data type knowledge. Kutikov et al. [29] established a urological label ontology project for the standardization of social media communication. Grandi [30] researched the personalized hierarchical management techniques based on multi-version ontology, and introduced a storage scheme that allows the representation and management of the relational database and a multi-version ontology evolution hierarchical structure. Kontopoulos et al. [31] proposed an ontology-based decision support tool and applied to the selection of solar hot water system optimization. Hoyle and Brass [32] raised the theory of statistical mechanics in the process of annotating objects whose annotated terminology comes from the ontology domain. Solano et al. [33] introduced ontology techniques for integrated processing and detection processes that use ontologies to focus on the evaluation of resource capabilities. Aime and Charlet [34] made use of the social psychology knowledge to ontology.

It can be seen that in the background of big data, the similarity formula of early heuristic method design has great limitation. In order to overcome the above problems and break through the constraints of the traditional ontology calculation model, the method of machine learning is widely applied to the ontology similarity calculation.

Since the goal of ontology learning is to obtain the similarity of each pair of vertices, the similarity of all vertex pairs is regarded as a symmetric similarity matrix, and the optimal ontology similarity matrix can be obtained by the strategy of matrix learning. The aim of this paper is to consider how to get a good similarity matrix or distance matrix. As a good distance matrix, it is measured from the following aspect: $\boldsymbol{M}$ should retain the structure of the ontology. This means that the distance between the vertexes should be smaller than the distance between the vertexes. Thus, the good distance matrix M should be able to effectively remove the noise during the dimensionality reduction process.

The organization of the rest paper is listed as follows: first we introduce the setting of our ontology problem; then the ontology distance matrix learning algorithm is presented in the section III; at last, we show its effectiveness in view of several experiments.

II. Setting

In this section, we introduce the setting of our ontology learning problem. In order to put the ontology algorithm into mathematical background, firstly we need to use d dimensional vector to express all the information of each vertex in ontology graph. This d dimensional vector includes the structure, name, attribute and instance information of its corresponding vector (concept). Hence, the ontology vertex space (instance space) V becomes a subset of $\mathbb{R}^d$. Throughout this paper, for the sake of convenience, v is used to represent the ontology of the vertex, the corresponding ontology concept of the vertex, and its corresponding d-dimensional vector. Readers can determine the meaning of the symbol according to the context, and the vector is no longer expressed in bold.

The Hadamard product $\otimes$ of two matrices with the same dimension is denoted by $(\boldsymbol{A}\otimes\boldsymbol{B})_{ij}=A_{ij}B_{ij}$, $\|\cdot\|_M$ denotes the Mahalanobis normal of matrix and $\|\cdot\|_F$ represents Frobenius norm which is defined as $\|\boldsymbol{A}\|_F=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|A_{ij}|^2}$. For any matrix $\boldsymbol{A}$, $\boldsymbol{A}\succ 0$ means $\boldsymbol{A}$ is a positive defined matrix. A function d is regarded as a distance function which satisfies the following properties (for any $v_i,v_j,v_k\in V$):

- non-negativity: $d(v_i,v_j)\ge 0$;
- symmetry: $d(v_i,v_j)=d(v_j,v_i)$;
- Triangle inequality: $d(v_i,v_j)+d(v_j,v_k)\ge d(v_i,v_k)$;
- Equivalence: $d(v_i,v_j)=0 \Leftrightarrow$ ontology vertices v_i and v_j corresponding to the same concept.

A Mahalanobis matrix $\boldsymbol{M}$ is a positive definite matrix, and the Mahalanobis distance between two ontology vertices $v_1,v_2\in V$ can be formulated as

$$d_{\boldsymbol{M}}^2(v_1,v_2)=\|v_1-v_2\|_M^2=(v_1-v_2)^{\mathrm{T}}M(v_1-v_2). \quad (1)$$

If $\boldsymbol{M}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$ and $v'=\boldsymbol{P}v$. Then we derive

$$d_{\boldsymbol{M}}^2(v_1,v_2)=(v_1-v_2)^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}(v_1-v_2)$$
$$=(\boldsymbol{P}v_1-\boldsymbol{P}v_2)^{\mathrm{T}}(\boldsymbol{P}v_1-\boldsymbol{P}v_2)=\|v_1^{'}-v_2^{'}\|^2.$$

The ontology distance matrix leaning problem can be summarized as:

$$\arg\min_{\boldsymbol{M}\succ 0} L(\boldsymbol{M})+r(\boldsymbol{M}),$$

where $L(\boldsymbol{M})$ denotes the loss part (or cost part) and $r(\boldsymbol{M})$ as the balance part.

Let Φ be a Riemannian manifold. The Stiefel manifold $Sm(p,d)$ is defined as the set of $d\times p$ ($p\le d$) matrices with orthonormal columns endowed with the Frobenius inner product constructing a compact Riemannian manifold:

$$Sm(p,d)=\{\boldsymbol{W}\in\mathbb{R}^{d\times p}:\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{I}_p\}.$$

The SPD manifold Γ_p is defined as the set of $p\times p$ dimensional real matrices endowed with the affine invariant Riemannian metric

$$\Gamma_p=\{\boldsymbol{M}\in\mathbb{R}^{p\times p}:\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x}>0,\forall\boldsymbol{x}\in\mathbb{R}^p-\{\boldsymbol{0}_p\}.$$

By a simple computation, we see that the dimensionality of $Sm(p,d)$ and Γ_p are $dp-\frac{p(p+1)}{2}$ and $\frac{p(p+1)}{2}$, respectively.

The aim of ontology learning algorithm becomes to a dimensionality reduction projection $\boldsymbol{W}:\mathbb{R}^d\to\mathbb{R}^d$ and a Mahalanobis matrix $\boldsymbol{M}\in\Gamma_p$. Let $S=\{(v_i,\tilde{v}_i,y_i)\}_{i=1}^n$ be the ontology training sample, where $v_i,\tilde{v}_i\in\mathbb{R}^p$ and

$$y_i=\begin{cases}1,\ v_i \text{ and } \tilde{v}_i \text{ are similar}\\ 0,\ v_i \text{ and } \tilde{v}_i \text{ are not similar}\end{cases}.$$

Hence, we have

$$d^2_{\boldsymbol{M},\boldsymbol{W}}(v_i,\tilde{v}_i)=(\boldsymbol{W}^{\mathrm{T}}v_i-\boldsymbol{W}^{\mathrm{T}}\tilde{v}_i)^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{W}^{\mathrm{T}}v_i-\boldsymbol{W}^{\mathrm{T}}\tilde{v}_i)$$
$$=(v_i-\tilde{v}_i)^{\mathrm{T}}\boldsymbol{W}\boldsymbol{M}\boldsymbol{W}^{\mathrm{T}}(v_i-\tilde{v}_i).$$

For $\alpha>0$, let

$$t_i=\exp(\alpha(v_i-\tilde{v}_i)^{\mathrm{T}}\boldsymbol{W}\boldsymbol{M}\boldsymbol{W}^{\mathrm{T}}(v_i-\tilde{v}_i)),$$

and then for pair of vertices $(v_i,\tilde{v}_i)$ with $y_i=1$, the loss is defined as

$$l(v_i,\tilde{v}_i\mid y_i=1)=\log(1+t_i);$$

for pair of vertices $(v_j,\tilde{v}_j)$ with $y_j=1$, the loss is defined as

$$l(v_j,\tilde{v}_j\mid y_j=0)=\log(1+t_j^{-1}).$$

In this way the ontology sample based cost function can be denoted as

$$L_S(\boldsymbol{W},\boldsymbol{M})=\sum_{i|y_i=1}\log(1+t_i)+\sum_{j|y_j=1}\log(1+t_j^{-1})$$
$$+\lambda r(\boldsymbol{M},\boldsymbol{M}_0),$$

where λ is a balance parameter and $r:\Gamma_p\times\Gamma_p\to\mathbb{R}^+$ is the regularizer, for example:

$$r(\boldsymbol{M},\boldsymbol{M}_0)=\mathrm{Tr}(\boldsymbol{M}\boldsymbol{M}_0^{-1})-\mathrm{logdet}(\boldsymbol{M}\boldsymbol{M}_0^{-1})-p.$$

Instead of setting the label y, we can divide the sample set into two parts S and D as follows:

$$S=\{(v_i,\tilde{v}_i)\mid v_i \text{ and } \tilde{v}_i \text{ shoud be similar}\},$$
$$D=\{(v_i,\tilde{v}_i)\mid v_i \text{ and } \tilde{v}_i \text{ shoud be dissimilar}\}.$$

More detailed, some papers even define the relative constraints as follows:

$$R=\{v_i,v_j,v_k\mid v_i \text{ should be more similar to } v_j \text{ than to } v_k\}.$$

Let F_Ω be the real-valued order-preserving function on the metric space Ω with distance metric d_Ω for a triplet $\delta_\Omega^{ijk}=(v_i,v_j,v_k)\in\Omega^3$ as

$$F_\Omega(\delta_\Omega^{ijk})=d_\Omega(v_i-v_j)-d_\Omega(v_i-v_k).$$

Then, the triplets in the ontology space V and real number response space Y are expressed as $\delta_V^{ijk}=(v_i,v_j,v_k)$ and $\delta_Y^{ijk}=(y_i,y_j,y_k)$. Moreover, the indicator function in this special setting can be stated as

$$\pi(i,j,k)=\begin{cases}1,\ \text{if } F_Y(\delta_Y^{ijk})<0\\ 1,\ \text{if } F_Y(\delta_Y^{ijk})=0 \text{ and } F_V(\delta_V^{ijk})<0\\ 0,\ \text{otherwise}\end{cases}.$$

In this way, the relative constraints are re-stated as follows:

$$R=\{v_i,v_j,v_k\mid\pi(i,j,k)=1\}.$$

Our ontology optimization problem becomes:

$$\min_{\boldsymbol{W},\boldsymbol{M}} L_S(\boldsymbol{W},\boldsymbol{M})$$
$$\text{s.t. } \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}=\boldsymbol{I}_p,\boldsymbol{M}\succ 0.$$

Let $\Theta_p=Sm(p,d)\times\Gamma_p$, and Υ_p be the orthogonal group. Then, for any $\boldsymbol{R}\in\Upsilon_p$, we have

$$L_S(\boldsymbol{W},\boldsymbol{M})=L_S(\boldsymbol{W}\boldsymbol{R},\boldsymbol{R}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{R})$$

and

$$\pi:\Theta_p\times\Upsilon_p\to\Theta_p:((\boldsymbol{W},\boldsymbol{M}),\boldsymbol{R})\to(\boldsymbol{W}\boldsymbol{R},\boldsymbol{R}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{R})$$

is the right group action on Θ_p.

III. ONTOLOGY DISTANCE MATRIX LEARNING

In this section, we present the main ontology distance matrix learning algorithm. First, we discuss the ontology learning model from multiple perspectives. Second, we present the main ontology learning algorithm in our paper which is proposed based on the coordinate descent and iterative tricks.

A. Manifold metric based ontology distance matrix and its kernelizing solution

By means of quotient manifold theory, the tangent space of Θ_p can be divided into two parts at $\Omega=(\boldsymbol{W},\boldsymbol{M})$: the vertical space $V_\Omega\Theta_p$ and horizontal space $H_\Omega\Theta_p$. Let g_p be the Riemannian metric of the product manifold Θ_p. Then, for any $v_\Omega\in V_\Omega\Theta_p$ and $h_\Omega\in H_\Omega\Theta_p$, we infer

$$g_p(h_\Omega,v_\Omega)=0.$$

An SPD matrix $\boldsymbol{M}\in\Gamma_p$ can be decomposed as $\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}}$, where $\boldsymbol{U}\in\Upsilon_p$ and $\boldsymbol{D}$ is a diagonal matrix with positive elements. With $\boldsymbol{W}\boldsymbol{U}=\boldsymbol{X}\in Sm(p,d)$, we deduce

$$\boldsymbol{W}\boldsymbol{M}\boldsymbol{W}^{\mathrm{T}}=\boldsymbol{W}\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}=\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}.$$

Therefore, our distance matrix ontology optimization problem can denote a cost term $L_S(\boldsymbol{X},\boldsymbol{D})$ with a search space defined by $Sm(p,d)\times\mathbb{R}_+^p$.

Now, we introduce how to kernelize the solution of our ontology learning problem. Let $\phi: V \to H$ be a mapping from ontology space V to a reproducing kernel Hilbert space (in short, RKHS) H with kernel function

$$K(v_i, v_j) = \langle \phi(v_i), \phi(v_j) \rangle.$$

For $\boldsymbol{M} \in \Gamma_p$ and $\boldsymbol{W} \in Sm(p, \dim(H))$, set

$$d_H^2(v_i, v_j) = (\phi(v_i) - \phi(v_j))^{\mathrm{T}} \boldsymbol{W}\boldsymbol{M}\boldsymbol{W}^{\mathrm{T}} (\phi(v_i) - \phi(v_j))$$

and

$$\tilde{t}_i = \exp(\alpha d_H^2(v_i, \tilde{v}_i)).$$

Then, we can re-write the ontology cost function as follows:

$$L_{S,H}(\boldsymbol{W}, \boldsymbol{M}) = \sum_{i|y_i=1} \log(1+\tilde{t}_i) + \sum_{j|y_j=1} \log(1+\tilde{t}_j^{-1})$$
$$+\lambda r(\boldsymbol{M}, \boldsymbol{M}_0).$$

By re-representing $\boldsymbol{W}$ as $\boldsymbol{W} = \Xi(\boldsymbol{D})\boldsymbol{A}$ with

$$\Xi(\boldsymbol{D}) = (\phi(d_1), \cdots, \phi(d_l)) \in \mathbb{R}^{\dim(H)\times l}$$

which denotes the matrix stacks the representation of l ontology training samples. Let $K(\boldsymbol{D}, \boldsymbol{D}) \in \Gamma_l$ be the kernel matrix with $[K(\boldsymbol{D}, \boldsymbol{D})]_{i,j} = K(d_i, d_j)$,

$$K^{\frac{1}{2}}(\boldsymbol{D}, \boldsymbol{D})\boldsymbol{A} = \boldsymbol{B} \in Sm(p, l)$$

such that $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{I}_p$, and

$$K(v, \boldsymbol{D}) = (K(v, d_1), \cdots, K(v, d_l))^{\mathrm{T}} \in \mathbb{R}^l.$$

In view of above definitions, we get

$$\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} = \boldsymbol{A}^{\mathrm{T}}\Xi(\boldsymbol{D})^{\mathrm{T}}\Xi(\boldsymbol{D})\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}K(\boldsymbol{D}, \boldsymbol{D})\boldsymbol{A} = \boldsymbol{I}_p,$$

and

$$d_H^2(v, \tilde{v}) = (\phi(v) - \phi(\tilde{v}))^{\mathrm{T}} \boldsymbol{W}\boldsymbol{M}\boldsymbol{W}^{\mathrm{T}} (\phi(v) - \phi(\tilde{v}))$$
$$= (K(v, \boldsymbol{D}) - K(\tilde{v}, \boldsymbol{D}))^{\mathrm{T}} K^{-\frac{1}{2}}(\boldsymbol{D}, \boldsymbol{D})\boldsymbol{B}\boldsymbol{M}\boldsymbol{B}^{\mathrm{T}}$$
$$\times K^{-\frac{1}{2}}(\boldsymbol{D}, \boldsymbol{D})(K(v, \boldsymbol{D}) - K(\tilde{v}, \boldsymbol{D})).$$

It implies that the ontology cost function can be formulated as a function of $\boldsymbol{B}$, i.e., $L_{S,H}(\boldsymbol{B}, \boldsymbol{M})$.

B. Ontology distance matrix learning iteration procedure

In this part, we mainly present the ontology distance matrix learning algorithm in the setting that Mahalanobis distance matrix is denoted by $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}}$ where $\boldsymbol{A} \in \mathbb{R}^{d\times n}$ ($r(\boldsymbol{A}) = n$) is certain linear transformation matrix and $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ ($n \le d$) is a diagonal matrix. Let $\boldsymbol{a}_i$ be the *i*-th column of ontology linear transformation matrix $\boldsymbol{A}$. Hence, the distance computation function (1) can be formulated as

$$d_{\boldsymbol{A},\boldsymbol{W}}^2(v_1, v_2) = (v_1 - v_2)^{\mathrm{T}} \boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}} (v_1 - v_2)$$
$$= (\boldsymbol{A}^{\mathrm{T}}(v_1 - v_2))^{\mathrm{T}} \boldsymbol{W} (\boldsymbol{A}^{\mathrm{T}}(v_1 - v_2))$$
$$= \sum_{i=1}^{m} W_{ii} (\langle v_1, \boldsymbol{a}_i \rangle - \langle v_2, \boldsymbol{a}_i \rangle)^2.$$

The ontology optimization problem can be denoted as

$$\arg\min_{\boldsymbol{W}\succ 0} \frac{\|\boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}}\|_F^2}{2} + C\sum_{i,j,k} \xi_{ijk}$$
$$\text{s.t. } \forall (v_i, v_j, v_k) \in R,$$
$$d_{\boldsymbol{A},\boldsymbol{W}}^2(v_i, v_k) - d_{\boldsymbol{A},\boldsymbol{W}}^2(v_i, v_j) \ge \varepsilon - \xi_{ijk},$$
$$\xi_{ijk} \ge 0. \qquad (2)$$

Here, ξ_{ijk} are slack variables and C is a positive constant. Let $\mathbf{w}$ be the diagonal vector of the matrix $\boldsymbol{W}$, then we have

$$d_{\boldsymbol{A},\boldsymbol{W}}^2(v_1, v_2) = \boldsymbol{w}^{\mathrm{T}}[(\boldsymbol{A}^{\mathrm{T}}v_1 - \boldsymbol{A}^{\mathrm{T}}v_2) \otimes (\boldsymbol{A}^{\mathrm{T}}v_1 - \boldsymbol{A}^{\mathrm{T}}v_2)].$$

Let

$$\boldsymbol{L} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) \otimes (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}),$$
$$\Delta_{v_1,v_2} = (\boldsymbol{A}^{\mathrm{T}}v_1 - \boldsymbol{A}^{\mathrm{T}}v_2) \otimes (\boldsymbol{A}^{\mathrm{T}}v_1 - \boldsymbol{A}^{\mathrm{T}}v_2),$$
$$z_{ijk} = \Delta_{v_i,v_k} - \Delta_{v_i,v_j}.$$

Then, the ontology optimization problem (2) can be stated as:

$$\arg\min_{\boldsymbol{w}} \frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{w}}{2} + C\sum_{i=1}^{n} \xi_n$$
$$\text{s.t. } \boldsymbol{w}^{\mathrm{T}} z_i \ge \varepsilon - \xi_i,$$
$$\xi_i, \boldsymbol{w}_j \ge 0. \qquad (3)$$

Since $\boldsymbol{L} \succ 0$, we ensure that ontology optimaztion problem (3) is a convex quadratic program.

Next, we define some positive parameters and multipliers: $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}$, and the Lagrangian function can be formulated by:

$$L(\boldsymbol{w}, \xi, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{w}}{2} + C\sum_{i=1}^{n} \xi_n$$
$$-\sum_{i=1}^{n} \lambda_i (\boldsymbol{w}^{\mathrm{T}} z_i - \varepsilon + \xi_i) - \sum_{i=1}^{n} \mu_i \xi_i - \sum_{i=1}^{n} \nu_i w_i.$$

To get the optimization solution of (3), we need to find the saddle point of $L(\boldsymbol{w}, \xi, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})$. By setting $\frac{\partial L(\boldsymbol{w}, \xi, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})}{\partial \boldsymbol{w}} = 0$, we obtain $\boldsymbol{w} = \boldsymbol{L}^{-1}(\sum_{i=1}^{n} \lambda_i z_i + \nu)$ and $C = \lambda_i + \mu_i$. Then, the ontology objective function is determined as:

$$-\frac{(\sum_{i=1}^{n} \lambda_i (z_i)^{\mathrm{T}})\boldsymbol{L}^{-1}(\sum_{i=1}^{n} \lambda_i (z_i))}{2} - \frac{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{L}^{-1}\boldsymbol{\nu}}{2}$$
$$-\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{L}^{-1}\sum_{i=1}^{n} \lambda_i (z_i) + \varepsilon \sum_{i=1}^{n} \lambda_i.$$

Set $\rho(\boldsymbol{\nu}) = \frac{\boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{L}^{-1}\boldsymbol{\nu}}{2}$, $\vartheta(\boldsymbol{\nu}, \lambda) = \boldsymbol{\nu}^{\mathrm{T}}\boldsymbol{L}^{-1}\sum_{i=1}^{n} \lambda_i (z_i)$, and $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ with $H_{ij} = z_i^{\mathrm{T}}\boldsymbol{L}^{-1}z_j$. The ontology learning problem (3) can be re-stated as

$$\arg\min_{\lambda,\nu} g(\lambda, \boldsymbol{\nu}) = \frac{\lambda^{\mathrm{T}}\boldsymbol{H}\lambda}{2} + \rho(\boldsymbol{\nu}) + \vartheta(\boldsymbol{\nu}, \lambda) - \varepsilon\sum_{i=1}^{n} \lambda_i$$

$$\text{s.t. } C \ge \lambda_i \ge 0, \nu_j \ge 0. \qquad (4)$$

Next, we present the coordinate descent trick to solve the ontology learning problem (4). Specifically, we describe an iterative algorithm to solve our ontology learning problem. Set (λ_0, ν_0) as the initial point, $\{(\lambda_k, \nu_k)\}_{k=0}^{\infty}$ and as the sequence of iterative vectors. Then, the detailed ontology algorithm is listed as follows:

Step 1. Mathematizing ontology information. For each vertex in ontology graph, we use a vector to express all its information. In the ontology mapping case, let $G_1, G_2, \dots, G_m$ be ontology graphs corresponding to ontologies $O_1, O_2, \dots, O_m$.

Step 2: Fix initial parameter of (λ_0, ν_0) and set

$$\boldsymbol{w} \leftarrow L^{-1}(\sum_{i=1}^{n} \lambda_i z_i + \nu);$$

$$\boldsymbol{T} \leftarrow L^{-1}\sum_{i=1}^{n} \lambda_i z_i;$$

Step 3: if (λ, ν) is not optimal, we do the following iteration:

- Randomly permute $(1, \cdots, n)$ to $(\pi(1), \cdots, \pi(n))$;

For $i = \pi(1), \cdots, \pi(n)$, $\Psi_\lambda \leftarrow (z_i)^{\mathrm{T}} \boldsymbol{w} - \varepsilon$,

$$\Lambda_\lambda = \begin{cases} \Psi_\lambda, & \text{if } 0<\lambda_i < C \\ \min\{0, \Psi_\lambda\}, & \text{if } \lambda_i = 0 \\ \max\{0, \Psi_\lambda\}, & \text{if } \lambda_i = C \end{cases}.$$

If $|\Lambda_\lambda| \ne 0$, then $\lambda_i^{\text{old}} \leftarrow \lambda_i$,

$$\lambda_i \leftarrow \min\{\max\{\lambda_i - \frac{\Psi_\lambda}{H_{ii}}, 0\}, C\},$$

$$\boldsymbol{w} \leftarrow \boldsymbol{w} + (\lambda_i - \lambda_i^{\text{old}})\boldsymbol{L}^{-1} z_i,$$

$$\boldsymbol{T} \leftarrow \boldsymbol{T} + (\lambda_i - \lambda_i^{\text{old}})\boldsymbol{L}^{-1} z_i.$$

And then $\nu^{\text{old}} \leftarrow \nu$, $\boldsymbol{w} \leftarrow \boldsymbol{w} + \boldsymbol{L}^{-1}(\lambda_i - \lambda_i^{\text{old}})$.

Step 4. By means of iteration algorithm and distance computing formulation, we map each pair of ontology vertices to the real line and thus determine the similarity of them in light of their distance.

Step 5. For each $v \in V(G)$, we obtain the similar vertices according to the distance and return the outcomes to the users. In the case of ontology mapping, for $v \in V(G_i)$, where $1 \le i \le m$, we obtain the similar vertices and return the outcome to the users.

C. Dual coordinate ascent approach and applied in ontology learning

As a supplement, we present the techniques of stochastic dual coordinate accent, and show how to apply it in ontology similarity measuring and ontology mapping. More result and related ontology learning algorithms can refer to [41], [42], [43] and [44].

The standard proximal stochastic dual coordinate ascent trick can be described as follows. In order to minimum (here l is an ontology loss function and Ω is a balance function which is used to control the sparsity of $\boldsymbol{w}$)

$$L(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n} l_i(\boldsymbol{V}_i^{\mathrm{T}}\boldsymbol{w}) + \lambda\Omega(\boldsymbol{w}),$$

we fixed the expected accuracy ε, initial dual solution $\alpha^{(0)} = 0$, $x^{(0)} = \frac{1}{\lambda n}\sum_{i=1}^{n} V_i \alpha_i^{(0)}$, $\boldsymbol{w}^{(0)} = \nabla\Omega^*(0)$. Let t be the iterative variable. Then, for each iteration, we randomly select i and determine $\Delta\alpha_i$ by means of the following method:

$$\Delta\alpha_i = \arg\max_{\Delta\alpha_i}\{-l_i^*(-(\alpha_i^{(t-1)} + \Delta\alpha_i)) - w^{(t-1)^{\mathrm{T}}} V_i \Delta\alpha_i - \frac{1}{2\lambda n}\|V_i\Delta\alpha_i\|^2\},$$

$$u = -\nabla l_i(V_i^{\mathrm{T}} w^{(t-1)}), q = u - \alpha_i^{(t-1)},$$

$$\Delta\alpha_i = \arg\max_{s\in[0,1]}\{-l_i^*(-(\alpha_i^{(t-1)} + sq)) - s w^{(t-1)^{\mathrm{T}}} V_i q - \frac{s^2}{2\lambda n}\|X_i q\|^2\},$$

$$s = \min\{1, \frac{l_i(V_i^{\mathrm{T}} w^{(t-1)}) + l_i^*(-\alpha_i^{(t-1)}) + w^{(t-1)^{\mathrm{T}}} V_i \alpha_i^{(t-1)} + \frac{\gamma}{2}\|q\|^2}{\|q\|^2(\gamma + \frac{1}{\lambda n}\|X_i\|^2)}\},$$

then reset $s = \frac{\lambda n\gamma}{R^2 + \lambda n\gamma}$ where R is the bound of $\|X_i\|$,

$$\alpha_i^{(t)} \leftarrow \alpha_i^{(t-1)} + \Delta\alpha_i,$$

$$\alpha_j^{(t)} \leftarrow \alpha_j^{(t-1)} \text{ if } j \ne i,$$

$$x^{(t)} \leftarrow x^{(t-1)} + (\lambda n)^{-1} V_i \Delta\alpha_i,$$

$$\boldsymbol{w}^{(t)} \leftarrow \nabla\Omega^*(x^{(t)}).$$

Let

$$T_0 = t - n - \left\lceil \frac{R^2}{\lambda_\gamma} \right\rceil < t,$$

$$\bar{\alpha} = \frac{\sum_{i=T_0+1}^{t} \alpha^{(i-1)}}{t - T_0},$$

$$\bar{\boldsymbol{w}} = \frac{\sum_{i=T_0+1}^{t} \boldsymbol{w}^{(i-1)}}{t - T_0}.$$

Let $\bar{\alpha} = \alpha^{(i)}$ and $\bar{\boldsymbol{w}} = \boldsymbol{w}^{(i)}$ for certain random $i \in T_0 + 1, \cdots, t$. If $L(\bar{\boldsymbol{w}}) - D(\bar{\alpha}) \le \varepsilon$, then we stop the iteration and output $\bar{\alpha}$ and $\bar{\boldsymbol{w}}$.

The corresponding accelerated process to minimum $L(\boldsymbol{w})$ can be stated as follows. Set

$$\kappa = \frac{R^2}{\gamma n} - \lambda, \mu = \frac{2}{\lambda}, \rho = \mu + \kappa, \eta = \sqrt{\frac{\mu}{\rho}}, \beta = \frac{1-\eta}{1+\eta},$$

$$\boldsymbol{y}^{(1)} = \boldsymbol{w}^{(1)} = 0, \alpha^{(1)} = 0, \xi_1 = (1 + \eta^{-2})(L(0) - D(0)).$$

For $t = 2, 3, \cdots$, the iterative process can be described as follows:

Set $\tilde{L}_t(w) = \frac{1}{n}\sum_{i=1}^{n} l_i(V_i^{\mathrm{T}}\boldsymbol{w}) + \tilde{\lambda}\tilde{\Omega}t(\boldsymbol{w})$,

where

$$\tilde{\lambda}\tilde{\Omega}t(\boldsymbol{w}) = \lambda\Omega(\boldsymbol{w}) + \frac{\kappa}{2}\|\boldsymbol{w}\|_2^2 - \kappa\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}^{(t-1)}.$$

Let $(w^{(t)}, \alpha^{(t)}, \varepsilon_t) = \mathrm{Prox}(\tilde{P}_t, \frac{\eta}{2(1+\eta^{-2})}\xi_{t-1}, \alpha^{(t-1)})$,

$$\boldsymbol{y}^{(t)} = \boldsymbol{w}^{(t)} + \beta(\boldsymbol{w}^{(t)} - \boldsymbol{w}^{(t-1)}),$$

$$\xi_t = (1-\eta/2)^{t-1}\xi_1.$$

If $t \ge 1 + \frac{2}{\eta}\log\frac{\xi_1}{\varepsilon}$ or

$$(1+\frac{\rho}{\mu}\varepsilon_t) + \frac{\rho\kappa}{2\mu}\left\|\boldsymbol{w}^{(t)} - \boldsymbol{y}^{(t-1)}\right\|^2 \le \varepsilon,$$

then the iterative process is stopped and $\boldsymbol{w}^{(t)}$ is returned.

Now, we discuss the details of the special case when ontology loss function is square loss and balance function is denoted as $\Omega(\boldsymbol{w}) = \frac{\lambda}{2}\|\boldsymbol{w}\|^2$, i.e.,

$$L(\boldsymbol{w}) = \frac{1}{2n}\sum_{i=1}^{n}(v_i^{\mathrm{T}}\boldsymbol{w} - \boldsymbol{y}_i)^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|^2.$$

In this case, we infer

$$\Delta\alpha_i = \arg\max_b -\frac{1}{2}(\alpha_i^{(t+1)} + b)^2 + \boldsymbol{y}_i(\alpha_i^{(t+1)} + b)$$

$$-\boldsymbol{w}^{(t-1)\mathrm{T}}x_i b - \frac{b^2\|v_i\|^2}{2\lambda n}$$

$$= \arg\max_b -\frac{1}{a}(1 + \frac{\|v_i\|^2}{2\lambda n})b^2 - (\alpha_i^{(t+1)} + \boldsymbol{w}^{(t-1)\mathrm{T}}x_i - y_i)b$$

$$= -\frac{\alpha_i^{(t+1)} + w^{(t-1)\mathrm{T}}v_i - y_i}{1 + \frac{\|v_i\|^2}{2\lambda n}}.$$

Thus, to minimize

$$L(\boldsymbol{w}) = \frac{1}{2n}\sum_{i=1}^{n}(v_i^{\mathrm{T}}\boldsymbol{w} - \boldsymbol{y}_i)^2 + \lambda(\frac{1}{2}\|\boldsymbol{w}\|^2 - \boldsymbol{w}^{\mathrm{T}}\mathrm{z}),$$

we use the following steps. Set

$$x^{(0)} = \frac{1}{\lambda n}\sum_{i=1}^{n}\alpha_i^{(0)}v_i,$$

$$\tilde{y}_i = y_i - v_i^{\mathrm{T}}z \text{ for any } i.$$

Randomly select i, and set

$$\Delta\alpha_i = -\frac{\alpha_i^{(t-1)} + x^{(t-1)\mathrm{T}}v_i - \tilde{y}_i}{1 + \frac{\|v_i\|^2}{2\lambda n}},$$

$$\alpha_i^{(t)} \leftarrow \alpha_i^{(t-1)} + \Delta\alpha_i,$$

$$\alpha_j^{(t)} \leftarrow \alpha_j^{(t-1)} \text{ for } j \ne i,$$

$$x^{(t)} \leftarrow x^{(t-1)} + \frac{\Delta\alpha_i}{\lambda n}v_i,$$

$$\boldsymbol{w}^{(t)} = x^{(t)} + z.$$

The whole process will be stopped if

$$\frac{1}{2n}\sum_{i=1}^{n}((v_i^{\mathrm{T}}\boldsymbol{w}^{(t)} - \boldsymbol{y}_i)^2 + (\alpha_i^{(t)} + \boldsymbol{y}_i)^2 - \boldsymbol{y}_i^2)$$

$$+\lambda\boldsymbol{w}^{(t)^{\mathrm{T}}}x^{(t)} \le \varepsilon.$$

The time complexity of the above implement is $\tilde{O}(d(n + \frac{R^2}{\lambda}))$, and it can be reduced to $\tilde{O}(d\sqrt{\frac{nR^2}{\lambda}})$ when $\frac{R^2}{\lambda} \gg n$.

Next, we consider the ontology learning setting when ontology loss is the logistic loss function, i.e.,

$$L(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n}\log(1 + e^{v_i^{\mathrm{T}}w}) + \frac{\lambda}{2}\|\boldsymbol{w}\|^2.$$

The corresponding ontology dual problem is stated as

$$D(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(\alpha_i\log(-\alpha_i) - (1+\alpha_i)\log(1+\alpha_i))$$

$$-\frac{\lambda}{2}\|x(\alpha)\|^2,$$

where $\alpha \in [0,1]^n$.

In order to minimize

$$L(w) = \frac{1}{n}\sum_{i=1}^{n}\log(1 + e^{v_i^{\mathrm{T}}w}) + \lambda(\frac{1}{2}\|\boldsymbol{w}\|^2 - \boldsymbol{w}^{\mathrm{T}}z),$$

we implement the following process. Set

$$x^{(0)} = \frac{1}{\lambda n}\sum_{i=1}^{n}\alpha_i^{(0)}v_i,$$

$$p_i = x_i^{\mathrm{T}}z \text{ for any } i.$$

Define $l^*(b) = b\log(b) + (1-b)\log(1-b)$. For each iteration,

$$p = v_i^{\mathrm{T}}w^{(t-1)},$$

$$q = -1/(1+e^{-p}) - \alpha_i^{(t-1)},$$

$$s = \min(1, \frac{\log(1+e^p) + l^*(-\alpha_i^{(t-1)}) + p\alpha_i^{(t-1)} + 2q^2}{q^2(4 + \frac{1}{\lambda n}\|v_i\|^2)}),$$

$$\Delta\alpha_i = sq,$$

$$\alpha_i^{(t)} = \alpha_i^{(t-1)} + \Delta\alpha_i,$$

$$\alpha_j^{(t)} = \alpha_j^{(t-1)} \text{ for } j \ne i,$$

$$x^{(t)} = x^{(t-1)} + \frac{\Delta\alpha_i}{\lambda n}v_i,$$

$$w^{(t)} = x^{(t)} + z,$$

The whole computation process halted if

$$\frac{1}{n}\sum_{i=1}^{n}(\log(1 + e^{v_i^{\mathrm{T}}w^{(t)}}) + l^*(-\alpha_i^{(t-1)})) + \lambda\boldsymbol{w}^{(t)\mathrm{T}}x^{(t)} \le \varepsilon.$$

Now, we discuss the Lasso case (parameter σ is a positive real number), i.e.,

$$\min\left[\frac{1}{2n}\sum_{i=1}^{n}(v_i^{\mathrm{T}}\boldsymbol{w}-\boldsymbol{y}_i)^2+\sigma\|\boldsymbol{w}\|_1\right].$$

By a simple computation, we deduce

$$\sigma\|\boldsymbol{w}\|_1\le\bar{y}\Rightarrow\|\bar{\boldsymbol{w}}\|_2\le\|\bar{\boldsymbol{w}}\|_1\le\frac{\bar{y}}{\sigma}.$$

The purpose in this special setting is to minimize the ontology learning frame

$$L(\boldsymbol{w})=\frac{1}{2n}\sum_{i=1}^{n}(v_i^{\mathrm{T}}\boldsymbol{w}-\boldsymbol{y}_i)^2+\lambda(\frac{1}{2}\|\boldsymbol{w}\|_2^2+\frac{\sigma}{\lambda}\|\boldsymbol{w}\|_1).$$

Note that

$$\left[\frac{1}{2n}\sum_{i=1}^{n}(v_i^{\mathrm{T}}\boldsymbol{w}^*-\boldsymbol{y}_i)^2+\sigma\|\boldsymbol{w}^*\|_1\right]$$

$$\le\left[\frac{1}{2n}\sum_{i=1}^{n}(v_i^{\mathrm{T}}\bar{\boldsymbol{w}}-\boldsymbol{y}_i)^2+\sigma\|\bar{\boldsymbol{w}}\|_1\right]+\frac{\lambda}{2}\|\bar{\boldsymbol{w}}\|_2^2+\frac{\varepsilon}{2}.$$

We have the following steps to minimize

$$L(\boldsymbol{w})=\frac{1}{2n}\sum_{i=1}^{\mathrm{n}}(v_i^{\mathrm{T}}\boldsymbol{w}-\boldsymbol{y}_i)^2$$

$$+\lambda(\frac{1}{2}\|\boldsymbol{w}\|^2+\sigma'\|\boldsymbol{w}\|_1-\boldsymbol{w}^{\mathrm{T}}z).$$

We set

$$x^{(0)}=\frac{1}{\lambda n}\sum_{i=1}^{n}\alpha_i^{(0)}v_i,$$

$$\boldsymbol{w}_j^{(0)}=\mathrm{sign}(x_j^{(0)}+z_j)\left[|x_j^{(0)}+z_j|-\sigma'\right]_+ \text{ for any } j.$$

In each iterative process, we randomly select i, and

$$\Delta\alpha_i=-\frac{\alpha_i^{(t-1)}+\boldsymbol{w}^{(t-1)\mathrm{T}}v_i-\boldsymbol{y}_i}{1+\frac{\|v_i\|^2}{2\lambda n}},$$

$$\alpha_i^{(t)}=\alpha_i^{(t-1)}+\Delta\alpha_i,$$

$$\alpha_j^{(t)}=\alpha_j^{(t-1)} \text{ for } j\ne i,$$

$$x^{(t)}=x^{(t-1)}+\frac{\Delta\alpha_i}{\lambda n}v_i,$$

$$\boldsymbol{w}_j^{(t)}=\mathrm{sign}(x_j^{(t)}+z_j)\left[|x_j^{(t)}+z_j|-\sigma'\right]_+ \text{ for any } j.$$

We stop the iteration if

$$\frac{1}{2n}\sum_{i=1}^{n}((v_i^{\mathrm{T}}\boldsymbol{w}^{(t)}-y_i)^2-2y_i\alpha_i^{(t)}+(\alpha_i^{(t)})^2)$$

$$+\lambda\boldsymbol{w}^{(t)\mathrm{T}}x^{(t)}\le\varepsilon.$$

Clearly, we can verify that the time complexity of the above process is $\tilde{O}(d(n+\min\left\{\frac{R^2}{\varepsilon\sigma^2},\sqrt{\frac{nR^2}{\varepsilon\sigma^2}}\right\}))$.

In addition, for the linear SVM setting, the ontology learning model can be formulated as

$$L(\boldsymbol{w})=\frac{1}{n}\sum_{i=1}^{n}\left[1-v_i^{\mathrm{T}}\boldsymbol{w}\right]_++\frac{\lambda}{2}\|\boldsymbol{w}\|^2.$$

For any a, we infer

$$l(a)-\frac{\gamma}{2}\le\tilde{l}(a)\le l(a).$$

Let $\tilde{L}$ be the SVM objective with smooth ontology hinge loss. Then, for any $\boldsymbol{w}$ and $\boldsymbol{w}'$, we get

$$L(\boldsymbol{w}')-L(\boldsymbol{w})\le\tilde{L}(\boldsymbol{w}')-\tilde{L}(\boldsymbol{w})+\frac{\gamma}{2}.$$

We discuss the ontology optimization problem with smooth ontology hinge loss, i.e.,

$$L(\boldsymbol{w})=\frac{1}{n}\sum_{i=1}^{n}l(v_i^{\mathrm{T}}\boldsymbol{w})+\lambda(\frac{1}{2}\|\boldsymbol{w}\|^2-\boldsymbol{w}^{\mathrm{T}}z).$$

Set

$$\boldsymbol{w}^{(0)}=z+\frac{1}{\lambda n}\sum_{i=1}^{n}\alpha_i^{(0)}v_i.$$

For each iteration, randomly select i,

$$\Delta\alpha_i=\max(-\alpha_i^{(t-1)},\min(1-\alpha_i^{(t-1)},\frac{1-v_i^{\mathrm{T}}\boldsymbol{w}^{(t-1)}-\gamma\alpha_i^{(t-1)}}{\|v_i\|^2/(\lambda n)+\gamma})),$$

$$\alpha_i^{(t)}\leftarrow\alpha_i^{(t-1)}+\Delta\alpha_i,$$

$$\alpha_j^{(t)}=\alpha_j^{(t-1)} \text{ for } j\ne i,$$

$$\boldsymbol{w}^{(t)}\leftarrow\boldsymbol{w}^{(t-1)}+\frac{\Delta\alpha_i}{\lambda n}v_i.$$

We stop the iterative process if

$$\frac{1}{n}\sum_{i=1}^{n}(l(v_i^{\mathrm{T}}\boldsymbol{w}^{(t)})-\alpha_i^{(t)}+\frac{\gamma}{2}(\alpha_i^{(t)})^2)$$

$$+\lambda\boldsymbol{w}^{(t)^{\mathrm{T}}}(\boldsymbol{w}^{(t)}-z)\le\varepsilon.$$

The time complexity of this produce is

$$\tilde{O}(d(n+\min\left\{\frac{R^2}{\gamma\lambda},\sqrt{\frac{nR^2}{\gamma\lambda}}\right\})).$$

D. The matrix version of coordinate ascent approach

The purpose of this subsection is to present the matrix version of ontology learning technology presented in the above subsection. Let $\boldsymbol{V}$ be the ontology information matrix, and $\boldsymbol{y}$ be the response vector. Accordingly, the ontology optimization framework can be stated as

$$\min_{\boldsymbol{w}\in\mathbb{R}^d}\frac{\|\boldsymbol{Vw}-\boldsymbol{y}\|^2}{2}+\lambda<1,\boldsymbol{w}>$$

$$\text{s.t. } \boldsymbol{w}\ge 0.$$

The standard program to solve the above ontology problem can be briefly described as follows:

Initialize $\boldsymbol{w}^{(0)}\leftarrow 0^d$, $\boldsymbol{r}^{(0)}\leftarrow\boldsymbol{y}$

For t=1,2,⋯T, do

$i\leftarrow$ get next coordinate;

$\delta\leftarrow\max\{-w_i^{(t-1)},\frac{<\boldsymbol{V}_i,\boldsymbol{r}^{(t-1)}>-\lambda}{\|\boldsymbol{V}_i\|^2}\}$;

$\boldsymbol{w}^{(t)}\leftarrow\boldsymbol{w}^{(t-1)}+\delta\boldsymbol{e}_i$;

$\boldsymbol{r}^{(t)}\leftarrow\boldsymbol{r}^{(t-1)}-\delta\boldsymbol{V}_i$;

Output: $\boldsymbol{w}^{(T)}$

Using the coordinate ascent approach, we present the new procedure as follows:

Initialize: $\boldsymbol{w}^{(0)} \leftarrow 0^d$; $\boldsymbol{r}^{(0)} \leftarrow \boldsymbol{y}$; $\boldsymbol{rr} \leftarrow \boldsymbol{r}^{(0)}$; $\boldsymbol{q}^{(0)} \leftarrow 0$; $\pi \leftarrow A(\boldsymbol{w}^{(0)})$

For t=1,2,⋯T, do

$\boldsymbol{rr} \leftarrow \boldsymbol{r}^{(t-1)}$; $\pi \leftarrow A(\boldsymbol{w}^{(t-1)})$; $\boldsymbol{q}^{(t-1)} \leftarrow 0$;

$i \leftarrow$ get next coordinate

If $\boldsymbol{q}^{(t-1)} \le \pi_i$ and $\boldsymbol{w}_i^{t-1} = 0$, then $\boldsymbol{w}^{(t)} \leftarrow \boldsymbol{w}^{(t-1)}$; $\boldsymbol{r}^{(t)} \leftarrow \boldsymbol{r}^{(t-1)}$ and $\boldsymbol{q}^{(t)} \leftarrow \boldsymbol{q}^{(t-1)}$

Otherwise:

$$\delta \leftarrow \max\{-w_i^{(t-1)}, \frac{<\boldsymbol{V}_i, \boldsymbol{r}^{(t-1)}> - \lambda}{\|\boldsymbol{V}_i\|^2}\};$$

$$\boldsymbol{w}^{(t)} \leftarrow \boldsymbol{w}^{(t-1)} + \delta \boldsymbol{e}_i ;$$

$$\boldsymbol{r}^{(t)} \leftarrow \boldsymbol{r}^{(t-1)} - \delta \boldsymbol{V}_i ;$$

$$\boldsymbol{q}^{(t)} \leftarrow \boldsymbol{q}^{(t-1)} - 2\delta <\boldsymbol{V}_i, \boldsymbol{r}^{(t-1)} - \boldsymbol{rr}> + \delta^2 \|\boldsymbol{V}_i\|^2 ;$$

Output: $\boldsymbol{w}^{(T)}$

Here the function A can be formulated as follows:

Initialize: $\pi \leftarrow 0^d$;

For $i = 1, \cdots, d$ do

$$g_i \leftarrow <\boldsymbol{V}_i, \boldsymbol{V}\boldsymbol{w} - \boldsymbol{y}> + \lambda ;$$

$$\pi_i \leftarrow \text{sign}(g_i) \frac{g_i^2}{\|\boldsymbol{V}_i\|^2} ;$$

Output π

IV. EXPERIMENTS

In this section, four simulation experiments on ontology measure and ontology mapping are manifested respectively, in which we mainly measure the effectiveness of main ontology learning algorithm. After the distance matrix is obtained, and the ontology similarity is then yielded via distance formula. To make comparisons as accurate as possible, the main algorithm was run in C++, in light of available LAPACK and BLAS libraries for operation computation and linear algebra calculating. The following four experiments are implemented on a multi-core CPU with a memory of 32GB.

A. Ontology similarity measure experiment on GO data

We use gene "GO" ontology O_1 in our first experiment, which is constructed in the website http://www. geneontology. The basic structure of O_1 is presented in Figure 1. We choose $P@N$ as the criterion to measure the quality of the experiment data. The ontology learning approaches raised in Huang et al. [35], Gao and Liang [36] and Gao and Gao [37] are applied to the "GO" ontology. Then after obtaining the average precision ratio by virtue of these three ontology learning algorithms, we compare the results with our algorithm, and parts of the compared data can refer to Table I.

When N=3, 5, 10 or 20, the result is compared with the precision ratios inferred by ontology learning algorithms proposed in Huang et al. [35], Gao and Liang [36] and Gao and Gao [37], the precision ratios derived from our distance matrix ontology learning algorithms are higher. Furthermore, the precision ratios reveal the tendency to increase apparently as N increases. In conclusion, our algorithms turn out to be much better and much more effective than those ontology learning tricks raised by Huang et al. [35], Gao and Liang [36] and Gao and Gao [37].

B. Ontology similarity measure experiment on PO data

We use O_2 , a plant "PO" ontology in our second experiment which was constructed in www.plantontology.org. The structure of O_2 presented in Fig. 2, and $P@N$ is used again to test the quality of the experiment data. Meanwhile, ontology methods in Huang et al. [35], Gao and Liang [36], and Wang et al. [38] are employed to the "PO" ontology. Then after yielding the average precision ratio in view of these three ontology learning algorithms, we compare the results with our ontology algorithm based on distance matrix learning. Parts of the compared data can be referred to Table II.

As seen in Table II, when N=3, 5 or 10, compared with the precision ratio determined by ontology learning algorithms raised in Huang et al. [35], Gao and Liang [36], and Wang et al. [38] the precision ratio got from our ontology algorithm is higher. Moreover, the precision ratios imply the tendency to increase apparently as N increases. As a result, our distance matrix ontology learning algorithms turn out to be much better and much more effective than those proposed by Huang et al. [35], Gao and Liang [36], and Wang et al. [38].

C. Ontology mapping experiment on humanoid robotics data

In our third experiment, we use "humanoid robotics" ontologies O_3 and O_4 to test the effectiveness of our proposed ontology learning algorithm. The structure of O_3 and O_4 are presented in Fig. 3 and Fig. 4 respectively, where ontology O_3 denotes the leg joint structure of bionic walking device for six-legged robot, and ontology O_4 shows the exoskeleton frame of a robot with wearable and power assisted lower extremities. The aim of experiment is to build the ontology mapping between O_3 and O_4 by means of similarity computation between the vertices of them. Again, we take $P@N$ precision ratio as a measure for the quality of experiment results. By applying ontology learning algorithms in Huang et al. [35], Gao and Liang [36], and Gao and Lan [39] on "humanoid robotics" ontologies and obtaining the average precision ratio, we compare the precision ratios got from these three approaches. Several compared results can refer to Table III.

Compared with the precision ratios (when N=1, 3 or 5) got by algorithms proposed in Huang et al. [35], Gao and Liang [36], and Gao and Lan [39], the precision ratios obtained from our distance matrix based ontology learning algorithm are higher. Furthermore, the precision ratios reveals the tendency to increase apparently as N increases. In all, our ontology learning algorithm turn out to be much better and much more effective than those raised by Huang et al. [35], Gao and Liang [36], and Gao and Lan [39] for the humanoid robotics engineering application.

D. Ontology mapping experiment on physics education data

In our last experiment, we use 'physics education' ontologies O_5 and O_6 to test the effectiveness of our proposed ontology learning algorithm with respect to ontology mapping. The ontology structures of O_5 and O_6 are represented in Fig. 5 and Fig. 6, respectively. The purpose of this experiment is to get ontology mapping between O_5 and O_6 by means of similarity computation. As before, we take $P@N$ precision ratio as a criterion for measuring the quality of the experiment results. To compare the results, we use ontology learning algorithms in Huang et al. [35], Gao and Liang [36] and Gao et al. [40] on "physics education" ontology. Parts of the compared results can refer to Table IV.

It can be seen that when N=1, 3 or 5, compared with the precision ratio obtained by ontology learning algorithms raised in Huang et al. [35], Gao and Liang [36] and Gao et al. [40], the precision ratio in terms of our distance matrix based ontology learning algorithm are much higher. Moreover, the precision ratios reveal the tendency to increase apparently as N increases. In conclusion, our ontology learning algorithms turn out to be much better and much more effective than those introduced by Huang et al. [35], Gao and Liang [36] and Gao et al. [40].

V. CONCLUSION

The goal of the ontology similarity computation algorithm is to obtain the similarity function through the learning of the ontology sample set, and then get the similarity of each pair of vertices. If the pairwise similarity is placed in a symmetric matrix, then it can be understood that the goal of the ontology learning algorithm is to get the similarity matrix, or the distance matrix (the shorter the distance is, the higher the similarity is; the smaller the distance is, the smaller the similarity is).

In this paper, we propose an ontology algorithm based on distance matrix learning. The approach to get the optimal solution is coordinate decent and iterative computation. Also, the manifold based ontology optimization framework is discussed and the kernelizing solution approach is presented. Four experiments have fully verified the feasibility of this proposed ontology algorithm in the specific application fields. The distance matrix learning algorithm provided in this paper has a wide range of applications in the field of ontology engineering.

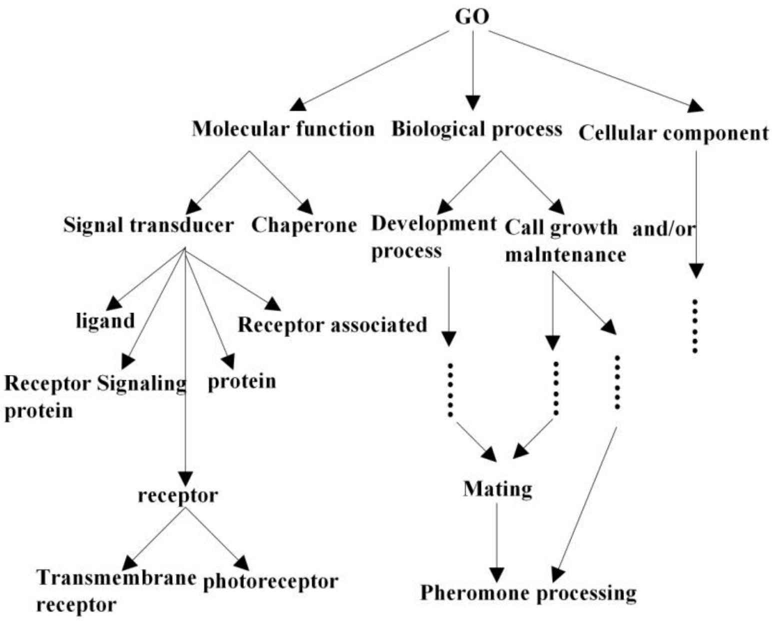

Fig. 1. The basic structure of GO ontology

TABLE I
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
Our Algorithm	0.4893	0.5764	0.6941	0.8672
Algorithm in [35]	0.4638	0.5348	0.6234	0.7459
Algorithm in [36]	0.4356	0.4938	0.5647	0.7194
Algorithm in [37]	0.4213	0.5183	0.6019	0.7239

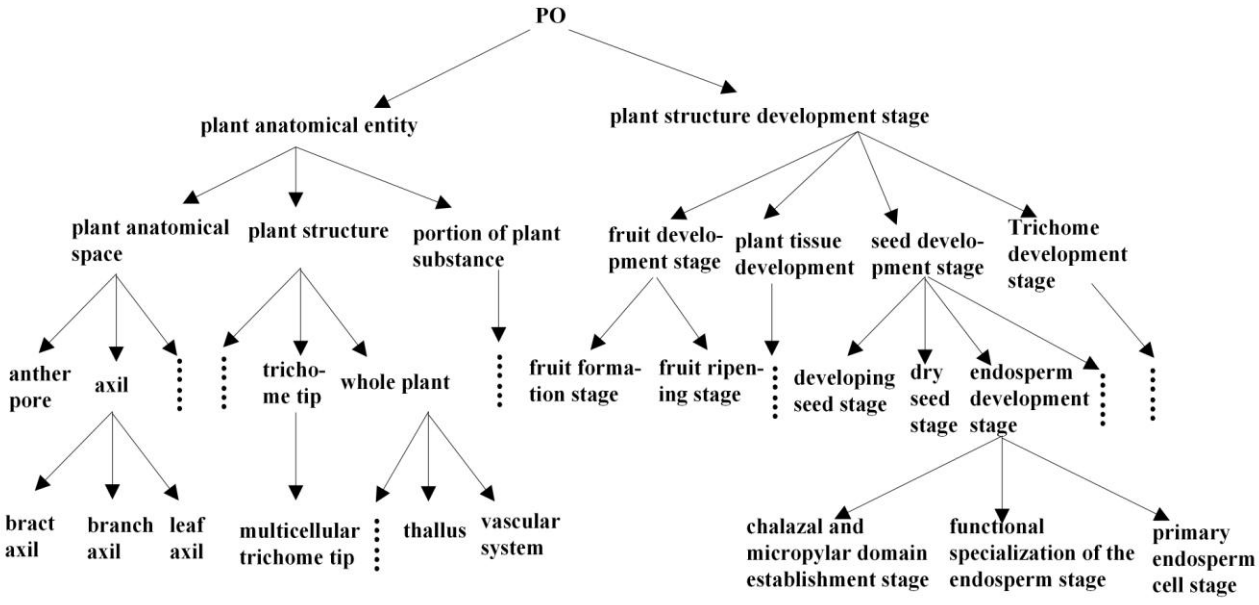

Fig. 2. The basic structure of PO ontology

TABLE II
THE EXPERIMENT RESULTS OF ONTOLOGY SIMILARITY MEASURE

	P@3 average precision ratio	*P*@5 average precision ratio	*P*@10 average precision ratio
Our Algorithm	0.4893	0.5607	0.7139
Algorithm in [35]	0.4282	0.4849	0.5632
Algorithm in [36]	0.4282	0.4849	0.5632
Algorithm in [38]	0.4549	0.5117	0.5859

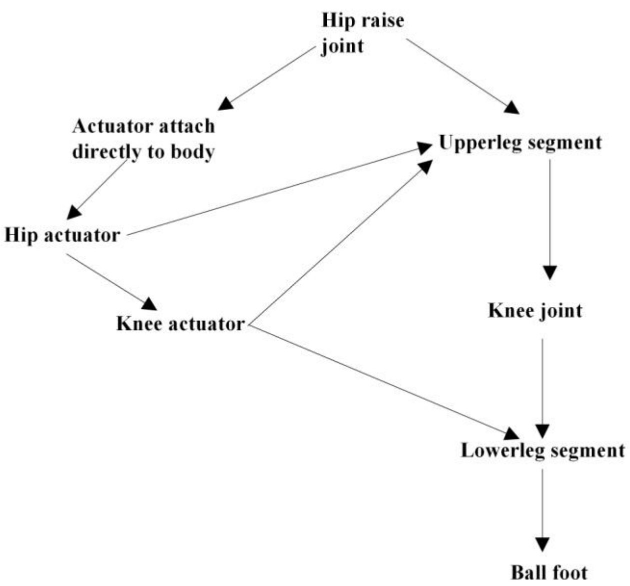

Fig. 3. 'Humanoid Robotics' Ontology O_3

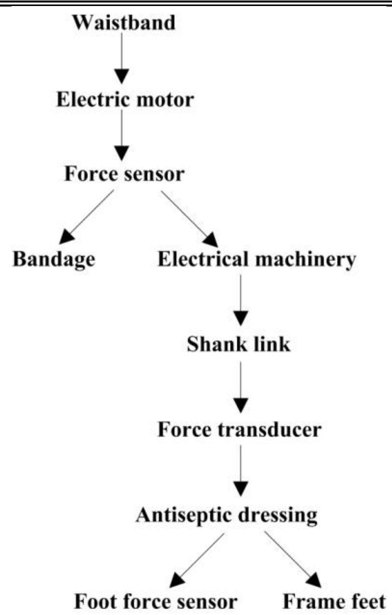

Fig. 4. 'Humanoid Robotics' Ontology O_4

TABLE III
THE EXPERIMENT RESULTS OF ONTOLOGY MAPPING

	P@1 average precision ratio	*P*@3 average precision ratio	*P*@5 average precision ratio
Our Algorithm	0.2778	0.5185	0.6778
Algorithm in [35]	0.2222	0.4074	0.4889
Algorithm in [36]	0.2778	0.4630	0.5333
Algorithm in [39]	0.2778	0.4815	0.5444

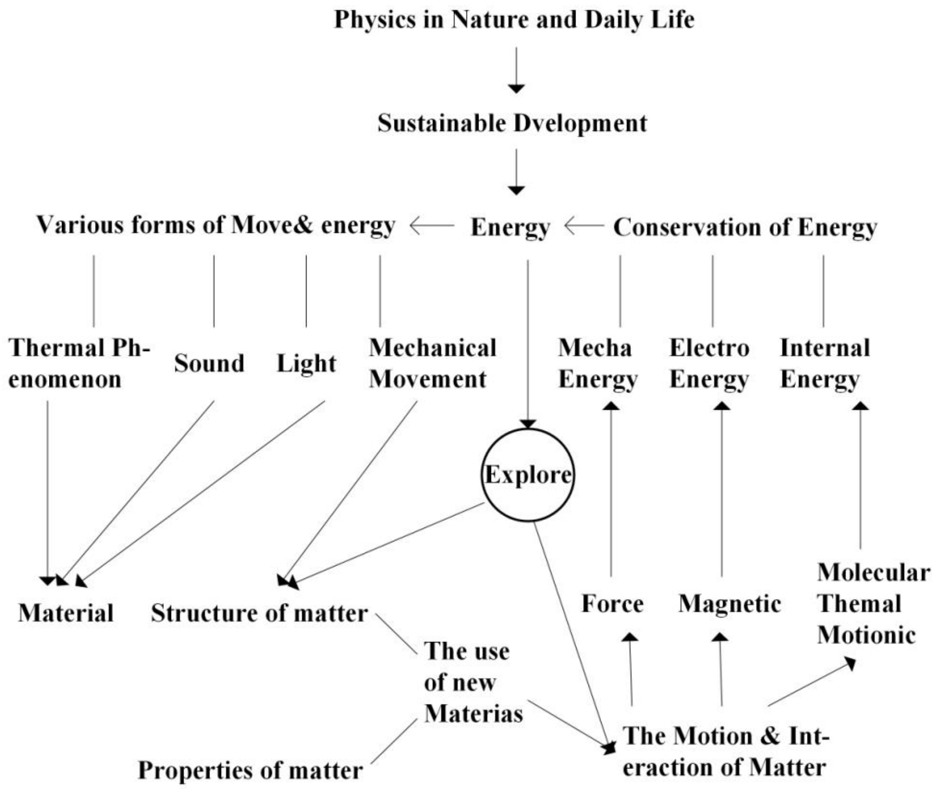

Fig. 5. 'physics education' Ontology O_5

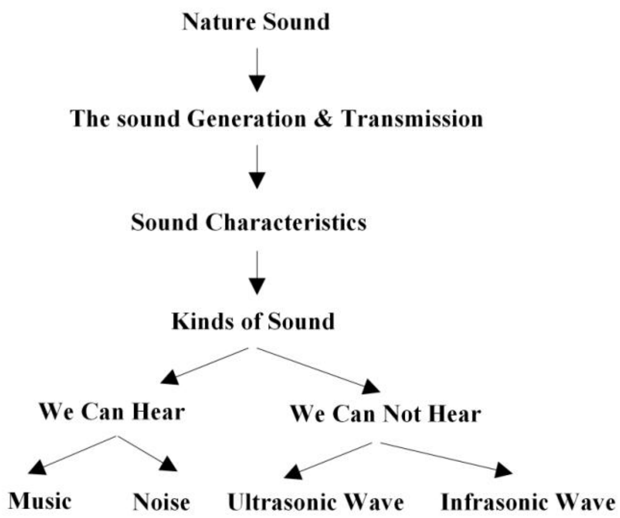

Fig. 6. 'physics education' Ontology O_6

TABLE IV
THE EXPERIMENT RESULTS OF ONTOLOGY MAPPING

	P@1 average precision ratio	*P*@3 average precision ratio	*P*@5 average precision ratio
Our Algorithm	0.6913	0.8065	0.9290
Algorithm in [35]	0.6129	0.7312	0.7935
Algorithm in [36]	0.6913	0.7556	0.8452
Algorithm in [40]	0.6774	0.7742	0.8968

.

REFERENCE

[1] P. Mork and P. Bernstein, "Adapting a generic match algorithm to align ontologies of human anatomy," *In: 20th International Conf. on Data Engineering*, Los Alamitos, CA, USA, Publisher: IEEE Comput. Soc. 2002, pp.787-790.

[2] F. Fonseca, M. Egenhofer, C. Davis, and G. Câmara, "Semantic granularity in ontology-driven geographic Information Systems," *AMAI Annals of Mathematics and Artificial Intelligence - Special Issue on Spatial and Temporal Granularity*, vol. 36, no. 1, pp. 121-151, 2002.

[3] A. M. Elizarov, A. B. Zhizhchenko, N. G. Zhil'tsov,A. V. Kirillovich, and E. K. Lipachev, "Mathematical knowledge ontologies and recommender systems for collections of documents in physics and mathematics," *Doklady Mathematics*, vol. 93, no. 2, pp. 231-233, 2016.

[4] N. M. Ochara, "Towards a regional ontology of management education in Africa: A complexity leadership theory perspective," *Acta Commercii*, vol. 17, no. 1, pp. 1-8, 2017.

[5] J. M. Przydzial, B. Bhhatarai, and A. Koleti, "GPCR ontology: development and application of a G protein-coupled receptor pharmacology knowledge framework," *Bioinformatics*, vol. 29, no. 24, pp. 3211-3219, 2013.

[6] S. Koehler, S. C. Doelken, and C. J. Mungall, "The human phenotype ontology project: linking molecular biology and disease through phenotype data," *Nucleic Acids Research*, vol. 42, no. D1, pp. 966-974, 2014.

[7] M. Ivanovic and Z. Budimac, "An overview of ontologies and data resources in medical domains," *Expert Systerms and Applications*, vol. 41, no. 11, pp. 5158-15166, 2014.

[8] A. Hristoskova, V. Sakkalis, and G. Zacharioudakis, "Ontology-driven monitoring of patient's vital signs enabling personalized medical detection and alert," *Sensors*, vol. 14, no. 1, pp. 1598-1628, 2014.

[9] M. A. Kabir, J. Han, and J. Yu, "User-centric social context information management: an ontology-based approach and platform," *Personal and Ubiquitous Computing*, vol. 18, no. 5, pp.1061-1083, 2014.

[10] Y. L. Ma, L. Liu, K. Lu, B. H. Jin, and X. J. Liu, "A graph derivation based approach for measuring and comparing structural semantics of ontologies", *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 5, pp. 1039-1052, 2014.

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

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

[13] T. Pizzuti, G. Mirabelli, M. A. Sanz-Bobi, F. Gomez-Gonzalez, "Food Track & Trace ontology for helping the food traceability control," *Journal of Food Engineering*, vol. 120, no. 1, pp. 17-30, 2014.

[14] N. Lasierra, A. Alesanco, and J. Garcia, "Designing an architecture for monitoring patients at home: Ontologies and web services for clinical and technical management integration," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 896-906, 2014.

[15] D. B. Carlini and M. Makowski, "Codon bias and gene ontology in holometabolous and hemimetabolous insects," *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, vol. 324, no. 8, pp. 686-698, 2015.

[16] C. Nicolai, "Deflationary truth and the ontology of expressions," *Synthese*, vol. 192, no. 12, pp. 4031-4055, 2015.

[17] C. G. P. Correae, A. M. de Carvalho Moura, and C. M. Claudia, "A multi-ontology approach to annotate scientific documents based on a modularization technique," *Journal of Biomedical Informatics*, vol. 58, pp. 208-219, 2015.

[18] H. A. Duran-Limon, C. A. Garcia-Rios, F. E. Castillo-Barrera, and R. Capilla, "An ontology-based product architecture derivation approach," *IEEE Transactions on Software Engineering*, vol. 41, no. 12, pp. 1153-1168, 2015.

[19] Y. Chabot, A. Bertaux, C. Nicolle, and T. Kechadi, "An ontology-based approach for the reconstruction and analysis of digital incidents timelines," *Digital Investigation*, vol. 15, pp. 83-100, 2015.

[20] J. P. Elbers and S. S. Taylor, "GO2TR: a gene ontology-based workflow to generate target regions for target enrichment experiments," Conservation Genetics Resources, vol. 7, no. 4, pp. 851-857, 2015.

[21] M. Rani, R. Nayak, and O. P. Vyas, "An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage," *Knowledge-Based Systems*, vol. 90, pp. 33-48, 2015.

[22] A. Sangiacomo, "The ontology of determination: from descartes to spinoza," *Sciencein Context*, vol. 28, no. 4, pp. 515-543, 2015.

[23] C. L. B. Azevedo, M. E. Iacob, J. P. A. Almeida, M. van Sinderen, L. F. Pires, and G. Guizzardi, "Modeling resources and capabilities in enterprise architecture: a well-founded ontology-based proposal for ArchiMate," *Information Systems*, vol. 54, pp. 235-262, 2015.

[24] H. Wimmer and R. Rada, "Good versus bad knowledge: ontology guided evolutionary algorithms," *Expert Systems with Applications*, vol. 42, no. 21, pp. 8039-8051, 2015.

[25] N. Trokanas and F. Cecelja, "Ontology evaluation for reuse in the domain of process systems engineering," *Computers & Chemical Engineering*, vol. 85, pp. 177-187, 2016.

[26] S. Chhun, Moalla, and Y. Ouzrout, "QoS ontology for service selection and reuse," *Journal of Intelligent Manufacturing*, vol. 27, no. 1, pp. 187-199, 2016.

[27] R. Costa, C. Lima, J. Sarraipa, and R. Jardim-Goncalves, "Facilitating knowledge sharing and reuse in building and construction domain: an ontology-based approach," *Journal of Intelligent Manufacturing*, vol. 27, no. 1, pp. 263-282, 2016.

[28] P. Panov, L. N. Soldatova, and S. Dzeroski, "Generic ontology of datatypes," *Information Sciences*, vol. 329, pp. 900-920, 2016.

[29] A. Kutikov, H. H. Woo, and J. W. Catto, "Urology tag ontology project: standardizing social media communication descriptors," *European Urology*, vol. 69, no. 2, pp. 183-185, 2016.

[30] F. Grandi, "Dynamic class hierarchy management for multi-version ontology-based personalization," *Journal of Computer and System Sciences*, vol. 82, no. 1, pp. 69-90, 2016.

[31] E. Kontopoulos, G. Martinopoulos, D. Lazarou, N. Bassiliades, "An ontology-based decision support tool for optimizing domestic solar hot water system selection," *Journal of Cleaner Production*, vol. 112, pp. 4636-4646, 2016.

[32] D. C. Hoyle and A. Brass, "Statistical mechanics of ontology based annotations," *Physica A-Statistical Mechanics and Its* Applications, vol. 442, pp. 284-299, 2016.

[33] L. Solano, F. Romero, and P. Rosado, "An ontology for integrated machining and inspection process planning focusing on resource capabilities," *International Journal of Computer Integrated Manufacturing*, vol. 29, no. 1, pp. 1-15, 2016.

[34] X. Aime and J. Charlet, "Social psychology insights into ontology engineering," *Future Generation Computer Systems-The International Journal of Escience*, vol. 54, pp. 348-351, 2016.

[35] X. Huang, T. W. Xu, W. Gao and Z. Y. Jia, "Ontology similarity measure and ontology mapping via fast ranking method," *International Journal of Applied Physics and Mathematics*, vol. 1, no. 1, pp. 54-59, 2011.

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

[37] Y. Gao and W. Gao, "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, 2012.

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

[39] W. Gao and M. H. Lan, "Ontology mapping algorithm based on ranking learning method," *Microelectronics and Computer*, vol. 28, no. 9, pp. 59-61, 2011.

[40] W. Gao, Y. Gao, and L. Liang, "Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping," *Journal of Chemical and Pharmaceutical Research*, vol. 5, no. 9, pp. 592-598, 2013.

[41] W. Gao, L. L. Zhu and Y. Guo, "Multi-dividing infinite push ontology algorithm," *Engineering Letters*, vol. 23, no. 3, pp. 132-139, 2015.

[42] M. H. Lan, J. Xu, and W. Gao, "Ontology feature extraction via vector learning algorithm and applied to similarity measuring and ontology mapping," *IAENG International Journal of Computer Science*, vol. 43, no.1, pp. 10-19, 2016.

[43] L. L. Zhu, W. G. Tao, X. Z. Min and W. Gao, "Theoretical characteristics of ontology learning algorithm in multi-dividing setting," *IAENG International Journal of Computer Science*, vol. 43, no.2, pp. 184-191, 2016.

[44] L. L. Zhu, Y. Pan, M. R. Farahani and W. Gao, "Theoretical characteristics on scoring function in multi-dividing setting," *IAENG International Journal of Applied Mathematics*, vol. 47, no. 1. Pp. 28-36, 2017.

Meihui Lan, female, was born in the city of Yiliang, Yunnan Province, China on Aug. 25, 1982. In 2006, she got bachelor degree in department of computer science and technology, Yunnan normal university, China. Then, she was enrolled in department of computer software and theory in the same school, and got Master degree and received a software designer certificate there in 2009.

Now, she acts as a lecturer in the department of computer science and engineering, Qujing Normal University. As a researcher in computer science, her interests are natural language processing and machine learning.

Jian Xu acts as an associate professor in the department of computer science and engineering, Qujing Normal University. As a researcher in computer science, his interests are Computer Education and Information Retrieval.

Wei Gao, male, was born in the city of Shaoxing, Zhejiang Province, China on Feb. 13, 1981. He got two bachelor degrees on computer science from Zhejiang industrial university in 2004 and mathematics education from College of Zhejiang education in 2006. Then, he was enrolled in department of computer science and information technology, Yunnan normal university, and got Master degree there in 2009. In 2012, he got PhD degree in department of Mathematics, Soochow University, China.

Now, he acts as an associate professor in the department of information, Yunnan Normal University. As a researcher in computer science and mathematics, his interests are covering two disciplines: Graph theory, Statistical learning theory, Information retrieval, and Artificial Intelligence.