Ontology Similarity Measuring and Ontology Mapping Algorithms Based on Proximal Technologies

Wei GAO[*,1], Jianzhang WU[2], Linli ZHU[3]

[1]*School of Information Science and Technology,* Yunnan Normal University, Kunming, Yunnan 650500, China
[2]*School of Computer Science and Engineer,* Southeast University, Nanjing 210096, China
[3]*School of Computer Engineering,* Jiangsu University of Technology, Changzhou 213001, China

Abstract — **In the field of information retrieval, ontology acts as an effective way to search the information that has highly semantic similarity of the original query concept, and return the results to the user. The relationship between different ontologies is created through ontology mapping, the essence of which is similarity computation. In this article, we present the sparse vector algorithms for ontology similarity measure and ontology mapping in terms of optimization frameworks. Proximal operator and proximal dividing technologies are used to determine the optimal solution. The simulation experimental results show that the new proposed algorithms have high efficiency and accuracy on ontology similarity measure and ontology mapping in multiple disciplines.**

Keywords - ontology; similarity measure; ontology mapping; sparse vector; proximal tricks

I. INTRODUCTION

Ontology is a knowledge representation and conceptual shared model which has been used in image retrieval, knowledge management and information retrieval search extension. As an effective concept semantic model, ontology is also used in other disciplines except computer science, including social science, medical science, biology science, pharmacology science and geography science (for instance, see Przydzial et al., [1], Koehler et al., [2], Ivanovic and Budimac [3], Hristoskova et al., [4], and Kabir et al., [5]).

In fact, the ontology model is a directed graph $G=(V,E)$, each vertex v in an ontology graph G stands for a concept and each directed edge $e=v_iv_j$ on an ontology graph G stands for a relationship between concepts v_i and v_j. The ontology similarity measure aims to find a similarity function Sim: $V\times V \to \mathbb{R}^{+} \cup \{0\}$ through which each pair of vertices is mapped to a real number. The ontology mapping aims 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 $v \in G_1$, determine a set $S_v \subseteq V(G_2)$ where the concepts is corresponding to vertices in S_v are semantic close to the concept correspond to v. There is one method to get such mapping, which is, for each $v \in G_1$, computing the similarity $S(v,v_j)$ where $v_j \in V(G_2)$ and choose a parameter $0<M<1$. Then S_v is a collection such that the element in S_v meets $S(v,v_j) \geq M$. In this way, the key of ontology mapping is to get a similarity function S and select a suitable parameter M.

In the past few years, ontology similarity-based technologies were used in various applications. On the basis of technology for stable semantic measurement, Ma et al., [6] presented a graph derivation representation based trick for stable semantic measurement. An ontology representation method was pointed by Li et al., [7], and it can be used for online shopping customers knowledge in enterprise information. Santodomingo et al., [8] reported a creative ontology matching system. The complex correspondences are presented by processing expert knowledge from external domain ontologies and in terms of using novel matching tricks. Pizzuti et al., [9] described the main features of the food ontology and some examples of application for traceability purposes. Lasierra et al., [10] proposed that ontologies are quite useful in designing an architecture for taking care of patients at home.

For ontology similarity measure and ontology mapping, there have several effective learning tricks. Wang et al., [11] 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., [12] presented fast ontology algorithm for calculating the ontology similarity in a short time. Gao and Liang [13] raised that the optimal ontology function can be determined by optimizing NDCG measure, and applied such idea in physics education. Gao and Gao [14] deduced the ontology function using the regression approach. Huang et al., [15] obtained ontology similarity function based on half transductive learning. Gao et al., [16] presented new ontology mapping algorithm using harmonic analysis and diffusion regularization on hypergraph. Gao and Shi [17] proposed new ontology similarity computation technology such that the new calculation model consider operational cost in the real implement. Very recently, Gao and Xu [18] presented the ontology similarity measuring and ontology mapping

DOI 10.5013/ IJSSST.a.17.43.01

ISSN: 1473-804x online, 1473-8031 print

algorithms based on minimum error entropy criterion. Several theoretical analysis of ontology algorithm cans refer to Gao et. al., [19], Gao and Xu [20], Gao and Zhu [21] and Gao et. al., [22].

In this paper, we present the new ontology similarity computation and ontology mapping algorithms based on sparse vector learning tricks. By means of the sparse vector, the ontology graph is mapped into a real line and vertices are mapped into real numbers. Then the similarity between vertices is measured by the difference between their corresponding real numbers. The rest of the paper is arranged as follows: we present the notations and setting in Section 2; two ontology optimization algorithms and iterative strategies are raised in Section 3; at last, the experiments on multiple disciples are designed to show the efficiency of the algorithm.

II. SETTING

Let V be instance space. For each vertex in ontology graph, its information (including its name, instance, attribute, structure, and semantic information of the concept which is corresponding to the vertex and that is contained in name and attribute components of its vector) is expressed by a p dimension vector. Let $v=\{v_1,\cdots,v_p\}$ be a vector corresponding to a vertex v. In order to facilitate the representation, we slightly confuse the notations and use v to denote both the ontology vertex and its corresponding vector. The ontology learning algorithms aim to get an optimal ontology (score) function f: $V\rightarrow\mathbb{R}$, the similarity between two vertices is judged by the difference between two real numbers which they correspond to. The essence of this algorithm is dimensionality reduction, i.e., use one dimension vector to express p dimension vector. Specifically, an ontology function f is a dimensionality reduction function f: $\mathbb{R}^p\rightarrow\mathbb{R}$.

Since the vector which corresponds to a vertex of ontology graph contains all the information of the vertex concept, attribute and the neighborhood structure in the ontology graph, it's always with high dimension. For instance, in the biological ontology, a vector may contain the information of all genes. In addition, ontology graph with large number of vertices makes ontology structure very complicated, and the most typical example is the GIS (Geographic Information System) ontology. These factors may led to the fact that the similarity calculation of ontology application is very large. But in fact, the similarity between the vertices is determined by small part of the vector components. For example, in the application of biological ontology, a genetic disease often results from a small number of genes, leaving most of the other genes irrelevant. Furthermore, in the application of geographic information system ontology, if an accident happens in a place and causes casualties, then we need to find the nearest hospital ignoring schools and shops nearby, i.e., we just need to find neighborhood information that meet specific requirements on the ontology graph. Therefore, tremendous academic and industrial interests are attracted to research the sparse ontology algorithm.

In the practice implement, one sparse ontology function is expressed by

$$f_{\beta}(v)=\sum_{i=1}^{p}v_i\beta_i+\delta. \quad (1)$$

Here $\boldsymbol{\beta}=(\beta_1,\cdots,\beta_p)$ is a sparse vector and δ is a noise term. The sparse vector $\boldsymbol{\beta}$ is to shrink irrelevant component to zero. To determine the ontology function f, we should learn the sparse vector $\boldsymbol{\beta}$ first.

One popular model with the penalize term via the l_1-norm of the unknown sparse vector $\boldsymbol{\beta}\in\mathbb{R}^p$:

$$Y_{\boldsymbol{\beta}} = l(\boldsymbol{\beta})+\lambda\|\boldsymbol{\beta}\|_1 \qquad \lambda>0 \quad . \quad (2)$$

Here, $\lambda > 0$ is a regularization parameter (or, balance parameter) and l is the principal function (in many articles, it called loss function) to measure the quality of $\boldsymbol{\beta}$. The balance term $\lambda\|\boldsymbol{\beta}\|_1$ is used to measure the sparsityof sparse vector $\boldsymbol{\beta}$. On the selection of the balance parameter λ, readers can refer to Mancinelli et al., [23], Zhu et al., [24], Mukhopadhyay and Bhattacharya [25], Ishibuchi and Nojima [26], Zhang et al., [27] and Varmuza et al., [28] for more details about the method of cross-validation. The target sparse vector is obtained by minimizing $Y_{\boldsymbol{\beta}}$.

A general structural ontology learning framework is given by:

$$Y_{\beta}=l(\boldsymbol{\beta})+\lambda g(\boldsymbol{\beta}) \quad (3)$$

where $l(\cdot)$ is a loss function, which is always supposed to be convex with respect to β, and $g(\cdot)$ is a convex balance function.

For $q\geq 1$, the l_q-norm of a vector $\boldsymbol{x}$ in $\mathbb{R}^n$ is denoted by $\|\boldsymbol{x}\|_q=(\sum_{i=1}^{n}|x_i|^q)^{1/q}$, where x_i is the i-th coordinate of v, and $\|\boldsymbol{x}\|_\infty=\lim_{q\to\infty}\|\boldsymbol{x}\|_q=\max_{i=1,\cdots,n}|\boldsymbol{v}_i|$. The l_0-pseudo-norm is defined by the number of nonzero components in a vector, i.e., $\|\boldsymbol{x}\|_0=\lim_{q\to 0^+}(\sum_{i=1}^{n}|x_i|^q)=|\{i:x_i\neq 0\}|$. Furthermore, the Frobenius norm of a matrix $\boldsymbol{V}\in\mathbb{R}^{m\times n}$ is denoted by $\|\boldsymbol{V}\|_F=(\sum_{i=1}^{m}\sum_{j=1}^{n}V_{ij}^2)^{\frac{1}{2}}$, where V_{ij} is the entry of

V in i-th row and j-th column. Let $(y)_+ = \max(y,0)$ for a scalar y.

We consider the solution of ontology problem (3) in the following setting: loss function l is differentiable with Lipschitz-continuous gradient, and g is a sum of l_∞-norms or l_2-norms since the l_∞-norm is piecewise linear by its definition.

III. MAIN ONTOLOGY ALGORITHMS

In this section, we present our main ontology sparse vector learning algorithms for ontology similarity measuring and ontology mapping. First, we manifest the standard proximal settings and the definition of proximal operator; then our model can be calculated via ontology flow problem, thus we propose our first ontology sparse vector learning algorithm based on the proximal operator and optimal of corresponding ontology flow problem in the third part; as a supplement of the first algorithm, the dual norm can be determined efficiently, which enables us to deduce the duality gaps of ontology optimization problem; in the last part of this section, we derive our second main ontology optimization algorithm by virtue of the proximal dividing technology.

A. Proximal Technology

In this part, the ontology optimization model is mainly followed by proximal technologies which have been widely used in Baingana et al., [29], Chao and Cheng [30], Xiao and Zhang [31], Beck and Teboulle [32], and Zhang et al., [33]. The gradient-based proximal tricks are stated by

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} l(\tilde{\boldsymbol{\beta}})+(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})\nabla l(\tilde{\boldsymbol{\beta}})+\lambda g(\boldsymbol{\beta}) \quad +\frac{L}{2}\left\|\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}\right\|_2^2 \tag{4}$$

here positive L is a upper bound on the Lipschitz constant of ∇l. Moreover, (4) can be rewritten as:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\left\|\boldsymbol{\beta}-\frac{1}{L}\nabla l(\tilde{\boldsymbol{\beta}})-\tilde{\boldsymbol{\beta}}\right\|_2^2+\frac{\lambda}{2}g(\boldsymbol{\beta}) \quad . \tag{5}$$

The proximal operator has been introduced by Moreau [34]. It is denote by $\text{prox}_{\lambda g}$ with balance term λg, which is the operator that maps a vector $\boldsymbol{u}\in\mathbb{R}^p$ to the unique solution of

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2}\left\|\boldsymbol{u}-\boldsymbol{\beta}\right\|_2^2+\lambda g(\boldsymbol{\beta}) \quad . \tag{6}$$

If $g(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$, then for each $i\in\{1,\cdots,p\}$, the proximal operator can be expressed as

$$u_i \to \text{sign}(\boldsymbol{u}_i)(|\boldsymbol{u}_i|-\lambda)_+ = \begin{cases} 0 & \text{if } |u_i|\le\lambda \\ \text{sign}(\boldsymbol{u}_i)(|\boldsymbol{u}_i|-\lambda) & \text{otherwise} \end{cases}.$$

If $g(\boldsymbol{u}) = \sum_{h\in\Omega}\|\boldsymbol{u}_h\|_2$ where Ω is a partition of $\{1,\cdots, p\}$, then for any $h\in\Omega$ the proximal operator can be denoted by

$$\boldsymbol{u}_h \to \boldsymbol{u}_h - \prod\nolimits_{\|\cdot\|_2\le\lambda}[\boldsymbol{u}_h] = \begin{cases} 0 & \text{if } \|\boldsymbol{u}_h\|_2\le\lambda \\ \dfrac{\|\boldsymbol{u}_h\|_2-\lambda}{\|\boldsymbol{u}_h\|_2}\boldsymbol{u}_h & \text{otherwise} \end{cases}$$

where $\prod_{\|\cdot\|_2\le\lambda}$ is the orthogonal map onto the λ- radius ball of the l_2-norm. Furthermore, if $g(\boldsymbol{u}) = \sum_{h\in\Omega}\|\boldsymbol{u}_h\|_\infty$, then for any $h\in\Omega$ the proximal operator can be similarly denoted by

$$\boldsymbol{u}_h \to u_h - \prod\nolimits_{\|\cdot\|_1\le\lambda}[\boldsymbol{u}_h]$$

where $\prod_{\|\cdot\|_1\le\lambda}$ is the orthogonal map onto the λ- radius l_1-ball.

Let $\xi = (\xi^h)_{h\in\Omega} \in \mathbb{R}^{p\times|\Omega|}$, and ξ_i^h be the i-th coordinate of ξ^g. Fixed $\boldsymbol{u}\in\mathbb{R}^p$, each solution $\xi^* = (\xi^{*h})_{h\in\Omega}$ of the optimal problem

$$\min_{\xi\in\mathbb{R}^{p\times|\Omega|}} \frac{1}{2}\left\|\boldsymbol{u}-\sum_{h\in\Omega}\xi^h\right\|_2^2 \text{ s.t. } \forall h\in\Omega,$$

$$\left\|\xi^h\right\|_1 \le \lambda\eta, \text{ and } \xi_i^h = 0 \text{ if } i \ne h. \tag{7}$$

meets $\boldsymbol{\beta}^* = \boldsymbol{u} - \sum_{h\in\Omega}\xi^{*h}$, where $\boldsymbol{\beta}^*$ is the solution of (6) such that g is a weighted sum of l_∞-norms. In terms of this fact, we focus on the model (7) for our ontology problem.

In what follows, we assume the entries of ξ and scalars $\boldsymbol{u}_j$ are non-negative. Then we associate our ontology problem with the ontology graph flow problem.

B. Ontology Graph Flow

For an ontology graph, we set the top vertex as a source vertex denoted by s, and add a sink vertex denoted by t which connects with all the leaf vertices in the ontology graph. Then the new ontology graph is denoted by $G'=(V,E, s, t)$. For all directed edges in E, we assign a non-negative capacity number, and a flow in ontology graph is exactly a non-negative function on directed edges such that capacity restrictions on all directed edges and conservation restrictions on all vertices except for the source and the sink. The real-valued cost function for each directed edge $e \in E$ relies on the value of the flow on e.

Let $\Psi \subseteq \{1,\cdots,p\}$ be a set of collections, and $(\eta_h)_{h\in\Psi}$ be positive weights. Let G=(V,E) be an ontology graph, its associated ontology graph G^* =(V,E, s, t) is uniquely denoted as follows: $V = V_u \cup V_{hr}$ where V_u is a vertex set with order p and one vertex being connected with each index $i \in \{1,\cdots,p\}$, and V_{hr} is a vertex set with order $|\Psi|$, one vertex per collection $h \in \Psi$. Hence, we have $|V| = |\Psi| + p$. In what follows, we identify indices $i \in \{1,\cdots, p\}$ and collections $h \in \Psi$ with vertices in the ontology graph, and thus we can refer to "vertex i" or "vertex h"; for each collection $h \in \Psi$, E includes a directed edge (s,h) and these directed edges with capacity $\lambda\eta_h$ and zero cost; for each collection $h \in \Psi$, and each index i in h, E includes a directed edge (h, i) with infinite capacity and zero cost, and we use ξ_i^h to denote the flow on this directed edge; for each index $i \in \{1,\cdots, p\}$, let $\bar{\xi}_i$ be the flow on (i, t), E contains a directed edge (i,t) cost $(\boldsymbol{u}_j - \bar{\xi}_j)^2/2$ and ∞ capacity.

As the flows ξ_j^h associated with G^*, the optimization problem (9) can be expressed as

$$\min_{\xi\in\mathbb{R}_+^{p\times|\Psi|},\bar{\xi}\in\mathbb{R}^p} \sum_{i=1}^{p}\frac{1}{2}(\boldsymbol{u}_i-\bar{\xi}_i)^2 \text{ s.t. } \bar{\xi} = \sum_{h\in\Psi}\xi^h \text{ and for any } h\in\Psi, \sum_{i\in h}\xi_i^h \le \lambda\eta_h \text{ and } \operatorname{supp}(\xi^h)\subseteq h . \quad (8)$$

Hence, finding a flow minimizing the sum of the costs on ontology associated graph is equivalent to solving problem (7).

C. Computation of Proximal Operator

Since $\bar{\xi} = \sum_{h\in\Psi}\xi^h$ is the only value for calculating the solution of $\boldsymbol{\beta} = \boldsymbol{u} - \bar{\boldsymbol{\xi}}$, we should first search the candidate value γ (it implies the lower bound $\frac{\|\boldsymbol{u}-\boldsymbol{\gamma}\|_2^2}{2}$ for the optimal cost of ontology problem (8).) for $\bar{\xi}$ by solving the relaxed version of problem (8) stated as follows:

$$\arg\min_{\gamma\in\mathbb{R}^p} \sum_{i\in V_u}\frac{1}{2}(u_i-\gamma_i)^2 \quad \text{s.t.} \quad \sum_{i\in V_u}\gamma_i \le \lambda\sum_{h\in V_{hr}}\eta_h . \quad (9)$$

Let F be the function for computing the flow. Our main algorithm then is stated below in view of the computation of proximal operator.

Algorithm 1. Calculating the proximal operator for collections.

Input $\boldsymbol{u} \in \mathbb{R}^p$, a set of collections Ψ, positive weights $(\eta_h)_{h\in\Psi}$, and positive balance parameter λ.

Step1. Constructing the initial ontology associated graph $G_0^* = (V_0, E_0, s, t)$.

Step 2. Calculating the optimal flow: $\bar{\xi} \leftarrow F(V_0, E_0)$ and return optimal solution of the proximal problem as $\boldsymbol{\beta} = u - \bar{\xi}$.

The function F(V=$V_u \cup V_{gr}$,E) in step 2 is described below:

Step 1. $\gamma \leftarrow \arg\min_{\gamma}\sum_{i\in V_u}\frac{1}{2}(u_i-\gamma_i)^2$

s.t. $\sum_{i\in V_u}\gamma_i \le \lambda\sum_{h\in V_{hr}}\eta_h$.

Step 2. For all vertices $i \in V_u$, set γ_i to be the capacity of the directed edge (i, t).

Step 3. Update $(\bar{\xi}_i)_{i\in V_u}$ by determining a max flow on the ontology associated graph (V,E, s, t).

Step 4. if there exists $i \in V_u$ s.t. $\bar{\xi}_i \ne \gamma_i$ then set (s, V^+) and (V^-, t) the two dis-joint subsets of (V, s, t) divided by the minimum (s, t)-cut of the ontology associated graph, and delete the directed edges between V^+ and V^-. Let E^+ and E^- be the rest two of disjoint subsets of E corresponding to V^+ and V^-. Then,

$(\bar{\xi}_i)_{i\in V_u^+} \leftarrow F(V^+, E^+)$,

$(\bar{\xi}_i)_{i\in V_u^-} \leftarrow F(V^-, E^-)$.

End if

Step 5. Return: $(\bar{\xi}_i)_{i\in V_u}$.

D. Determine the Dual Norm

For any vector $\kappa \in \mathbb{R}^p$, the dual norm g^* of g is defined by

$$g^*(\kappa) = \max_{g(z)\le 1} z^{\mathrm{T}}\boldsymbol{\kappa}.$$

Let l^* be the Fenchel conjugate of loss function l which is stated by $l^*(\boldsymbol{\kappa}) = \sup_z [z^{\mathrm{T}}\boldsymbol{\kappa} - l(z)]$. The duality gap for ontology problem (3) can be obtained by the standard Fenchel duality theory (see Liu and Liu [35], Bot and Heinrich [36], Sun [37], and Wang et al., [38] for more details) and it is equal to

$$l(\boldsymbol{\beta}) + \lambda\Omega(\boldsymbol{\beta}) + l^*(-\boldsymbol{\kappa}).$$

Thus, we should determine efficiently g^* for evaluating the duality gap and find a feasible dual variable $\boldsymbol{\kappa}$. This boils down to discussing another ontology flow problem, relied on the following variational expression:

$g^*(\boldsymbol{\kappa}) = \min_{\xi \in \mathbb{R}^{p\times|\Psi|}} \tau$ s.t. $\sum_{h\in\Psi} \xi^h = \boldsymbol{\kappa}$, and for any $h \in \psi$, we have

$$\left\|\xi^h\right\|_1 \le \tau\eta_h \text{ with } \xi_i^h = 0 \text{ if } i \notin h \qquad (10)$$

In the ontology problem (10), the capacities on the directed edges (s,h), $h \in \psi$, are set to $\tau\eta_h$, and the capacities on the directed edges (i, t), $i \in \{1,\cdots,p\}$, are fixed to $\boldsymbol{\kappa}_i$. We need to search the smallest value of t to solve problem (10), such that there exists a flow beginning all the capacities $\boldsymbol{\kappa}_i$ on the directed edges leading to the sink t. Let *DN* be the function to get the dual norm. Then, we describe the algorithm for determining the dual norm.

Algorithm 2. Calculating the dual norm.

Input $\boldsymbol{\kappa} \in \mathbb{R}^p$, a set of collections Ψ, positive weights $(\eta_h)_{h\in\Psi}$.

Step 1. Constructing the initial ontology associated graph $G_0^* = (V_0, E_0, s, t)$

Step 2. $\tau \leftarrow DN(V_0, E_0)$.

Step 3: Return value of the dual norm τ.

Here function *DN* is presented as follows:

DN(V $= V_u \cup V_{gr}$, *E*)

Step 1. $\tau \leftarrow \dfrac{\sum_{j\in V_u} \kappa_j}{\sum_{h\in V_{hr}} \eta_h}$ and set the capacities of directed edges (s,h) to $\tau\eta_h$ for any $h \in V_{hr}$.

Step 2. Update $(\overline{\xi}_i)_{i\in V_u}$ by computing a max-flow on the ontology associated graph (V, E, s, t).

Step 3. If there exists $i \in V_u$, s.t. $\overline{\xi}_i \ne \kappa_i$ then set (V^+, E^+) and (V^-, E^-) as in Algorithm 1, and assign $\tau \leftarrow DN(V^-, E^-)$.

Step 4. Return: τ.

This algorithm give the support of algorithm 1 such that the duality gaps for ontology optimization model can be determined efficiently.

E. Proximal Tricks for ontology Optimization

We now present proximal dividing methods for our ontology problem (3) which gives us another way to optimal the ontology sparse learning problem. Assume that l can be expressed as $l(\boldsymbol{\beta}) = \sum_{i=1}^n \tilde{l}_i(\boldsymbol{\beta})$, where the functions $\tilde{l}_i$ satisfies that $\mathrm{prox}_{\gamma\tilde{l}_i}$ can be deduced in closed form for any positive γ and any i—that is, for any $\boldsymbol{u}$ in $\mathbb{R}^m$, the following problems allows closed form solutions: $\min_{v\in\mathbb{R}^m} \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{x}\|_2^2 + \gamma\tilde{l}_i(\boldsymbol{x})$; or l can be formulated as $l(\boldsymbol{\beta}) = \tilde{l}(\boldsymbol{V\beta})$ for any $\boldsymbol{\beta} \in \mathbb{R}^p$.

For each collection $h \in \Psi$, we introduce additional variables $\boldsymbol{z}^h \in \mathbb{R}^{|\psi|}$. Thus, (3) can be further expressed as

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p, z^h\in\mathbb{R}^{|\Psi|} \text{ for } h\in\Psi} l(\boldsymbol{\beta}) + \lambda\sum_{h\in\Psi}\eta_h\left\|z^h\right\|$$

$$\text{s.t.} \quad \forall h \in \Psi \quad , \quad \boldsymbol{z}^h = \boldsymbol{\beta}_h \quad . \qquad (11)$$

To solve this problem, we use the technology of alternating direction (see Chen et al., [39], Boglaev [40], Sulc et al., [41], Bnouhachem and Ansari [42] and Tay and Tan [43] for more details). For each $h \in \Psi$, $x^h \in \mathbb{R}^{|\Psi|}$ is the dual variables. Assume augmented Lagrangian

$$\Xi(\boldsymbol{\beta}, (\boldsymbol{z}^h)_{h\in\Psi}, (v^h)_{h\in\Psi}) = l(\boldsymbol{\beta}) +$$

$$\sum_{h\in\Psi}[\lambda\eta_h\|\boldsymbol{z}\|^h + (\boldsymbol{v}^h)^{\mathrm{T}}(\boldsymbol{z}^h - \boldsymbol{\beta}_h) + \frac{\gamma}{2}\|\mathrm{z}^h - \boldsymbol{\beta}_h\|_2^2]$$

is dual variables with positive balance parameter γ. To solve Equation (11), we need to search a saddle point of the augmented Lagrangian, and it can be implemented by iterative computation stated as follows:

Algorithm 3. Ontology sparse vector learning based on proximal dividing

Step 1. Minimize Ξ regarding to $\boldsymbol{\beta}$ and other variables should keep fixed.

Step 2. Minimize Ξ regarding to $\boldsymbol{z}^h$, and other variables should keep fixed. The solution can be obtained in closed form: for any $h \in \Psi$, $\boldsymbol{z}^h \leftarrow \mathrm{prox}_{\frac{\lambda\eta_h}{\gamma}\|\cdot\|}[\boldsymbol{\beta}_h - \frac{1}{\gamma}\boldsymbol{v}^h]$.

Step 3. Take a gradient ascent step on Ξ regarding with the $\boldsymbol{v}^h$: $\boldsymbol{v}^h \leftarrow \boldsymbol{v}^h + \gamma(\boldsymbol{z}^h - \boldsymbol{\beta}_h)$.

Step 4. go back to step 1.

IV. EXPERIMENTS

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

In the following four experiment, we verify the effectiveness of main ontology algorithms (Algorithm 1 and Algorithm 3) in our paper. We check the availability of Algorithm 1 in our first and second experiments, and inspect Algorithm 3 in our third and fourth experiments. After getting the sparse vector $\boldsymbol{\beta}$, the ontology function then is derived by $f_\beta(v) = \sum_{i=1}^{p} v_i\beta_i$.

A. Experiment on Biology Data

"Go" ontology O_1 which was constructed in http: //www. geneontology. org. is used for our experiment. We use *P@N* (Precision Ratio, see Craswell and Hawking [44] for more detail) to measure the equality of the experiment. At first, the closest *N* concepts for every vertex on the ontology graph are to be given by experts, and then we get the first *N* concepts for every vertex on ontology graph by the Algorithm 1 and compute the precision ratio.

Ontology algorithms in Huang et al., [12], Gao and Liang [13] and Gao and Gao [14] are applied to "Go" ontology. In the end, we compare the precision ratio which we have obtained from the four methods. Several experiment results can be referred to Tab. 1.

TABLE 1. THE EXPERIMENT DATA FOR ONTOLOGY SIMILAIRTY MEASURE

	***P@*3 average precision ratio**	***P@*5 average precision ratio**	***P@*10 average precision ratio**	***P@*20 average precision ratio**
Algorithm 1	47.43%	56.16%	64.75%	85.74%
Algorithm in [12]	46.38%	53.48%	62.34%	74.59%
Algorithm in [13]	43.56%	49.38%	56.47%	71.94%
Algorithm in [14]	42.13%	51.83%	60.19%	72.39%

As *N*= 3, 5, 10 or 20, the precision ratio by means of Algorithm 1 is higher than that determined by algorithms which has been pointed out in Huang et al., [12], Gao and Liang [13] and Gao and Gao [14]. Particularly, such precision ratios are increasing obviously along with *N* increasing. Hence, the Algorithm 1 used in our paper is superior to the method presented by Huang et al., [12], Gao and Liang [13] and Gao and Gao [14].

B. Experiment on Physical Education Data

Physical education ontologies O_2 and O_3 (the structures of O_2 and O_3 can refer to [13] for more details) are chosen for our second experiment. This experiment aims at determining the ontology mapping between O_2 and O_3 via sparse ontology Algorithm 2. *P@N* criterion is applied to measure the equality of the experiment. At first, with the help of experts, the closest *N* concepts for each vertex on the ontology graph are obtained. Then we obtain the first *N* concepts for every vertex on ontology graph by means of the Algorithm 1 and compute the precision ratio.

At the same time, ontology algorithms in Huang et al., [12], Gao and Liang [13] and Gao et al., [16] are applied to "physical education" ontology. In the end, we compare the precision ratio that we have got from four methods. Several experiment results can be referred to Tab. 2.

TABLE 2. THE EXPERIMENT DATA FOR ONTOLOGY MAPPING

	P@1 average precision ratio	**P@3 average precision ratio**	**P@5 average precision ratio**
Algorithm 1	70.97%	81.72%	92.90%
Algorithm in [12]	61.29%	73.12%	79.35%
Algorithm in [13]	69.13%	75.56%	84.52%
Algorithm in [16]	67.74%	77.42%	89.68%

From the experiment results in Table 2, we find it more efficient to use Algorithm 1 than algorithms proposed in Huang et al., [12], Gao and Liang [13] and Gao et al., [16], especially in the situation where N is sufficiently large.

C. *Experiment on Plant Data*

In this subsection, we use "PO" ontology O_4 which was constructed in http: //www.plantontology.org. To test the efficiency of Algorithm 3 for ontology similarity measuring. Similarly, we choose the *P@N* standard once more for this experiment. Moreover, the ontology method in Wang et al., [11], Huang et al., [12] and Gao and Liang [13] is employed to the "PO" ontology. In the end, we compute the accuracy by the three algorithms and compare the results with Algorithm 3. Part of the data can be referred to Table 3.

TABLE 3. THE EXPERIMENT DATA FOR ONTOLOGY SIMILAIRTY MEASURE

	P@3 average precision ratio	**P@5 average precision ratio**	**P@10 average precision ratio**
Algorithm 3	48.66%	57.80%	72.82%
Algorithm in [11]	45.49%	51.17%	58.59%
Algorithm in [12]	42.82%	48.49%	56.32%
Algorithm in [13]	48.31%	56.35%	68.71%

As N= 3, 5, or 10, the precision ratio in terms of Algorithm 3 is higher than that determined by algorithms which has been raised in Wang et al., [11], Huang et al., [12] and Gao and Liang [13]. Particularly, such precision ratios are increasing obviously along with N increasing. Hence, the Algorithm 3 is superior to the method that has been raised by Wang et al., [11], Huang et al., [12] and Gao and Liang [13].

D. *Experiment on University Data*

We choose "University" ontologies O_5 and O_6 for our last experiment, and the structures of O_5 and O_6 can refer to [45] for more details. This experiment aims at determining ontology mapping between O_5 and O_6 via our sparse ontology Algorithm 3. We use *P@N* criterion to measure the equality of the experiment. Ontology algorithms in Wang et al., [11], Huang et al., [12] and Gao and Liang [13] are chosen to be applied to "University" ontologies. In the end, we compare the precision ratio which we have got from four methods. Several experiment results can be referred to Table 4.

TABLE 4. THE EXPERIMENT DATA FOR ONTOLOGY MAPPING

	P@1 average precision ratio	**P@3 average precision ratio**	**P@5 average precision ratio**
Algorithm 3	39.29%	59.52%	75.71%
Algorithm in [11]	28.57%	51.19%	60.71%
Algorithm in [12]	28.57%	42.86%	48.57%
Algorithm in [13]	32.14%	54.76%	65.00%

From the experiment results in Table 4, we find it more efficient to use Algorithm 3 than algorithms proposed in Wang et al., [11], Huang et al., [12] and Gao and Liang [13] especially in the situation where N is sufficiently large.

V. CONCLUSIONS

As a data structural representation and storage model, ontology has been widely used in various fields and proved to have high efficiency. The core of ontology algorithms is obtaining the similarity measure between vertices on given ontology graph. In recent years, various learning methods have been employed for ontology similarity measure and ontology mapping. One learning approach is mapping each vertex to a real number, and the similarity is judged by the difference between the real numbers which the vertices correspond to. The sparse ontology algorithm is suitable for ontology computation with big data and arouses the great concern among the researchers.

In this paper, a new computation technology is presented for ontology similarity measure and ontology mapping application. The tricks are based on the proximal operator and proximal dividing technology. At last, simulation data shows that our new algorithm has high efficiency in biology, physics education, plant science and university applications. The ontology sparse algorithm raised in our paper illustrates the promising application prospects for multiple disciplines.

ACKNOWLEDGEMENTS

We thank the reviewers for their constructive comments in improving the quality of this paper. The research is financed by: NSFC (No.11401519).

REFERENCES

[1] 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.

[2] 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.

[3] M. Ivanovic and Z. Budimac, "An overview of ontologies and data resources in medical domains", Expert Systerms and Applications, vol. 41, no. 11, pp. 5158-5166, 2014.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] T. Pizzuti, G. Mirabelli, M. A. Sanz-Bobi, and 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.

[10] 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.

[11] 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.

[12] 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, pp. 54-59, 2011.

[13] 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.

[14] 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.

[15] X. Huang, T. W. Xu, W. Gao, and S. Gong, "Ontology similarity measure and ontology mapping using half transductive ranking," In Proceedings of 2011 4th IEEE international conference on computer science and Information technology, Chengdu, China, 2011, pp. 571-574.

[16] 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.

[17] W. Gao and L. Shi, "Ontology similarity measure algorithm with operational cost and application in biology science", BioTechnology: An Indian Journal, vol. 8, no. 11, pp. 1572-1577, 2013.

[18] W. Gao and T. W. Xu, "Ontology similarity measuring and ontology mapping algorithm based on MEE criterion", Energy Education Science and Technology Part A: Energy Science and Research, vol. 32, no. 5, pp. 3793-3806, 2014

[19] W. Gao, Y. Gao, and Y. G. Zhang, "Strong and weak stability of k-partite ranking algorithm", Information, vol. 15, no. 11A, pp. 4585-4590, 2012.

[20] W. Gao and T. W. Xu, "Stability analysis of learning algorithms for ontology similarity computation", Abstract and Applied Analysis, 2013, 9 pages, http://dx.doi.org/10.1155/2013/174802.

[21] W. Gao and L. L. Zhu, "Gradient learning algorithms for ontology computing", Computational Intelligence and Neuroscience, vol. 2014, 12 pages, http://dx.doi.org/10.1155/2014/438291.

[22] W. Gao, L. Yan, and L. Liang, "Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (I)", International Journal of Computer Applications in Technology, vol. 50, no. (3/4), pp. 226-231, 2014.

[23] G. Mancinelli, S. Vizzini, A. Mazzola, S. Maci, and A. Basset, "Cross-validation of delta N-15 and FishBase estimates of fish trophic position in a Mediterranean lagoon: The importance of the isotopic baseline", Estuaring Coastal and Shelf Science, vol. 135, pp. 77-85, 2013.

[24] Z. Zhu, P. Chen, and J. Zhuang, "Predicting Chinese children and youth's energy expenditure using ActiGraph accelerometers: a calibration and cross-validation study", Research Quarterly for Exercise and Sport, vol. 84, no. 2 pp. S56-S63, 2013.

[25] S. Mukhopadhyay and S. Bhattacharya, "Cross-validation based assessment of a new Bayesian palaeoclimate model", Environmetrics, vol. 24, no. 8, pp. 550-568, 2013.

[26] H. Ishibuchi and Y. Nojima, "Repeated double cross-validation for choosing a single solution in evolutionary multi-objective fuzzy classifier design", Knowledge-based Systems, vol. 54, pp. 22-31, 2013.

[27] Y. Zhang, X. Yu, D. Guo, Y. Yin, and Z. Zhang, "Weights and structure determination of multiple-input feed-forward neural network activated by Chebyshev polynomials of Class 2 via cross-validation", Neural Computing & Applications, vol. 25, pp. 1761-1770, 2014 .

[28] K. Varmuza, P. Filzmoser, M. Hilchenbach, H. Krüger, and J. Silén, "KNN classification- evaluated by repeated double cross validation: Recognition of minerals relevant for comet dust", Chemometrics and Intelligent Laboratory System, vol. 138, pp. 64-71, 2014.

[29] B. Baingana, G. Mateos, and G. B. Giannakis, "Proximal-gradient algorithms for tracking cascades over social networks", IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 4, pp. 563- 575, 2014.

[30] M. T. Chao and C. Z. Cheng, "A note on the convergence of alternating proximal gradient method", Applied Mathematics and Computation, vol. 228, pp. 258-263, 2014

[31] L. Xiao and T. Zhang, "Proximal stochastic gradient method with progressive variance reduction", SIAM Journal on Optimization, vol. 24, no. 4, pp. 2057-2075, 2014.

[32] A. Beck and M. Teboulle, "A fast dual proximal gradient algorithm for convex minimization and applications", Operations Research Letters, vol. 42, no. 1, pp. 1-6, 2014.

[33] H. B. Zhang, J. Wei, M. X. Li, J. Zhou, and M. T. Chao, "On proximal gradient method for the convex problems regularized with the group reproducing kernel norm", Journal of Global Optimization, vol. 58, no. 1, pp. 169-188, 2014.

[34] J. J. Moreau. "Fonctions convexes duales et points proximaux dans un espace Hilbertien", Compte Rendus de lAcademie des Sciencs, Paris, Serie A, Mathematiques, vol. 255, pp. 2897–2899, 1962.

[35] J. W. Liu and Y. Liu, "Non-integer norm regularization SVM via Legendre-Fenchel duality", Neurocomputing, vol. 144, pp. 537-545, 2014.

[36] R. I. Bot and A. Heinrich, "Regression tasks in machine learning via Fenchel duality", Annals of Operations Research, vol. 222, no. 1, pp. 197-211, 2014.

[37] X. K. Sun, "Regularity conditions characterizing Fenchel-Lagrange duality and Farkas-type results in DC infinite programming", Journal of Mathematical Analysis and Applications, vol. 414, no. 2, pp. 590-611, 2014.

[38] S. S. Wang, Y. Xia, Q. G. Liu, P. Dong, D. D. Feng, and J. H. Luo, "Fenchel duality based dictionary learning for restoration of noisy images", IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 5214-5225, 2014.

[39] C. Chen, Ng, K. Michael, and X. L. Zhao, "Alternating direction method of multipliers for nonlinear image restoration problems", IEEE Transactions on Image Procssing, vol. 24, no. 1, pp. 33-43, 2015.

[40] I. Boglaev, "A uniform monotone alternating direction scheme for nonlinear singularly perturbed parabolic problems", Journal of Computational and Applied Mathematics, vol. 272, pp. 148-161, 2014.

[41] P. Sulc, S. Backhaus, and M. Chertkov, "Optimal distributed control of reactive power via the alternating direction method of multipliers", IEEE Transactions on Energy Conversion, vol. 29, no. 4, pp. 968-977, 2014.

[42] A. Bnouhachem and Q. H. Ansari, "A descent LQP alternating direction method for solving variational inequality problems with separable structure", Applied Mathematics and Computation, vol. 246, pp. 519-532, 2014.

[43] W. C. Tay and E. L. Tan, "Pentadiagonal alternating-direction-implicit finite-difference time-domain method for two-dimensional Schrodinger equation", Computer Physics Communications, vol. 185, no. 7, pp. 1886-1892, 2014.

[44] N. Craswell and D. Hawking, "Overview of the TREC 2003 web track", Proceeding of the Twelfth Text Retrieval Conference, Gaithersburg, Maryland, NIST Special Publication, 2003, pp. 78-92.

[45] M. H. Lan and W. Gao, "Ontology mapping based on k-partite ranking learning method", Journal of Suzhou University of Science and Technology (Natural Science), vol. 29, no. 2, pp. 60-62, 2012.

.