Ontology Sparse Vector Learning Based on Accelerated First-Order Method

Yun Gao¹ and Wei Gao²*

¹Department of Editorial, Yunnan Normal University, Kunming 650092, China; ²School of Information and Technology, Yunnan Normal University, Kunming 650500, China

Abstract: In this article, we present a sparse vector learning algorithm for ontology similarity measure and ontology mapping by virtue of accelerated first-order technology. The main procedure of our iterative algorithm is relying on proximity operator computation, Picard-Opial process and accelerated first-order tricks. The simulation experimental results show that the new proposed algorithm has high efficiency and accuracy in ontology similarity measure and ontology mapping in plant science and university application.

Keywords: Accelerated first-order method, ontology mapping, ontology, similarity measure, sparse vector.

1. INTRODUCTION

Ontology, a knowledge representation and conceptual shared model, has proved itself to be useful in image retrieval, knowledge management and information retrieval search extension. What’s more, as an effective concept semantic model, ontology also finds its place in the other disciplines like 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]).

Actually, the ontology model is a graph $G=(V,E)$, each vertex $v$ in an ontology graph $G$ stands for a concept and each edge $e=vw$ of an ontology graph $G$ stands for a relationship between concepts $v$ and $w$. Ontology similarity-based technologies were employed in a variety of applications several years ago. On the condition of technology for stable semantic measurement, a graph derivation representation which is based on trick for stable semantic measurement is proposed by Ma et al., [6]. Later, Li et al., [7] contributed to pointing out an ontology representation method which can be used for online shopping customers knowledge in enterprise information. Santotdomingo et al., [8] explained a creative ontology matching system. In his view, 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] gave a detailed description about the main features of the food ontology and some examples of application for traceability purposes. Lasiera et al., [10] proposed that ontologies are very useful in designing an architecture for looking after patients at home.

Concerning ontology similarity measure and ontology mapping, several effective learning tricks work well. Wang et al., [11] tended to learn a score function to map each vertex to a real number. Then, according to the difference of the real number which the two vertices correspond to, we can measure the similarity between them. Huang et al., [12] worked out a fast ontology algorithm to calculate the ontology similarity within a short time. Gao and Liang [13] reported that the optimal ontology function can be determined by optimizing NDCG measure. And they also took the idea to physics education. Gao and Gao [14] deduced the ontology function through the regression approach. Moreover, based on half transductive learning, Huang et al., [15] obtained ontology similarity function. Gao and Zhu [21] raised new ontology mapping algorithm by means of harmonic analysis and diffusion regularization on hypergraph. Gao and Shi [17] proposed new ontology similarity computation technology. As a result, the new calculation model considering operational cost in the real implement. Few years ago, Gao and Xu [18] presented the ontology similarity measuring and ontology mapping algorithms on basis of minimum error entropy criterion. Several theoretical analysis of ontology algorithm can 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 accelerated first-order based ontology algorithm for ontology similarity computation and ontology mapping. 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; the ontology optimization algorithm and iterative strategies are raised in Section 3; at last, the experiments on plant ontology and university ontology are designed to show the efficiency of the algorithm.

2. SETTING

Let $V$ be a instance space. For each vertex in ontology graph, a $p$ dimension vector expresses information including
its name, instance, attribute and structure, and semantic information of the concept which corresponds to the vertex and that is contained in name and attributes components of its vector. Let \( v = \{v_1, \cdots, v_p\} \) be a vector that corresponds to a vertex \( v \). To promote the representation, we try confusing the notations and using \( v \) to denote both the ontology vertex and its corresponding vector. The ontology learning algorithms are set to get an optimal ontology (score) function \( f: V \rightarrow \mathbb{R} \), and the similarity between two vertices is judged by the difference between two corresponding real numbers. The core of this algorithm is dimensionality reduction, i.e., choosing one dimension vector to express \( p \) dimension vector. In specific, an ontology function \( f \) is a dimensionality reduction function \( f: \mathbb{R}^p \oplus \mathbb{R} \).

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

\[
f_p(v) = \sum_{i=1}^{p} v_i \beta_i + \delta
\]

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

In our paper, we consider the general versions for learning \( \beta \). Let \( \{v_i, y_i\}_{i=1}^{n} \) be a sample set with \( n \) vertex, \( V = \oplus V_{V^n} \) be the matrix of \( n \) samples such that each sample vertex lies in a \( p \) dimension space, and \( y = (y_1, \cdots, y_n) \in \oplus \) be the vector of outputs of the these \( n \) sample vertex. Hence, the regression function (1) can be expressed as the linear model:

\[
y = V\beta + \delta
\]

where \( \delta \) is the \( n \) dimension vector for noise distributed as \( N(0, \sigma^2 I_{n \times n}) \).

The general regression obtains an estimate of the sparse vector by solving the following optimization problem:

\[
\min_{\beta \in \mathbb{R}^p} l(\beta) + \lambda \|\beta\|_1
\]

where \( l(\beta) = \frac{1}{2}\|y - V\beta\|_2^2 \) is the loss term, \( \|\beta\|_1 = \sum_{i=1}^{p} |\beta_i| \) is the \( l_1 \)-norm of sparse vector \( \beta \), and \( \lambda > 0 \) is the balance parameter which controls the sparsity level.

In what follows, let \( |\hat{\beta}| = (|\beta_i|)_{i=1} \in \mathbb{R}^p \) be the vector. For each \( q \geq 1 \), the \( l_q \)-norm of sparse vector \( \beta \) is denoted by

\[
\|\beta\|_q = \left( \sum_{i=1}^{p} |\beta_i|^{q} \right)^{\frac{1}{q}}.
\]

Let \( \delta_C: \mathbb{R}^p \oplus \mathbb{R} \) be the indicator function of the set \( C \subseteq \mathbb{R}^p \) such that \( \delta_C(x) = 0 \) if \( x \in C \) and \( \delta_C(x) = +\infty \) otherwise. Furthermore, let \( \partial \varphi(x) \) be the sub-differential of \( \varphi \) at \( x \in \mathbb{R}^p \) where \( \varphi: \mathbb{R}^p \oplus \mathbb{R} \) is a convex function.

### 3. MAIN ONTOLOGY ALGORITHM AND SOLUTION TECHNOLOGIES

#### 3.1. Optimization Framework

Let \( \rho > 0 \) be a parameter, \( \Lambda \) be a prescribed convex subset, and \( \Gamma: \mathbb{R}^p \times \mathbb{R} \rightarrow \mathbb{R} \) be a function given by

\[
\Gamma(\beta, \lambda) = \frac{1}{2} \sum_{i=1}^{p} \left( \frac{\beta_i^2}{\lambda_i} + \lambda_i \right).
\]

Then, our ontology optimization model can be expressed as

\[
\inf \left\{ \frac{1}{2} \|y - V\beta - y\|_2^2 + \rho \Gamma(\beta, \lambda) : \beta \in \mathbb{R}^p, \lambda \in \Lambda \right\}
\]

where the balance term takes the form

\[
\Omega(\beta) = \inf \Gamma(\beta, \lambda) : \lambda \in \Lambda \right\}.
\]

Since \( \Gamma \) is jointly convex, the framework (5) is still convex. Moreover, the balance function \( \Omega \) is independent of the signs of the elements of ontology sparse vector \( \beta \).

Now, we introduce proximity operator for optimizing our ontology model. Let \( \varphi \) be a real-valued convex function on \( \mathbb{R}^d \). For each \( t \in \mathbb{R}^d \), the proximity operator of \( \varphi \) is denoted by \( \text{prox}_\varphi(t) = \arg \min \frac{1}{2} \|z - t\|_2^2 + \varphi(z) : z \in \mathbb{R}^d \).

This operator is well-defined since the above minimum exists and unique.

#### 3.2. Selection of Generic Convex Set \( \Lambda \)

Now, we focus on the generic convex set \( \Lambda \). For each \( \lambda \in \mathbb{R}^+ \), the quadratic function \( \Gamma(\cdot, \lambda) \) supplies an upper bound to the \( l_1 \)-norm in view of \( \Omega(\beta) \geq \|\beta\|_1 \), and the equality is hold if and only if \( |\beta| \in \Lambda \) according to the arithmetic-geometric inequality.

For each solution \( (\beta, \lambda) \), the sparsity pattern of \( \beta \) is included in the sparsity pattern of \( \lambda \), i.e., the indices associated with nonzero components of \( \beta \) are a subset of those of \( \lambda \). If \( \lambda = 0 \), then \( \beta = 0 \) also holds in view of the ratio \( \beta_i^2 / \lambda_i \).
Thus, if the set $\Lambda$ supports some sparse solutions of $\hat{\lambda}$, then the same sparsity pattern will be manifested on $\hat{\beta}$. Furthermore, the sparse vectors $\lambda$ is supported by $\sum_i \lambda_i$. For instance, a restriction of the form $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ supports consecutive zeros at the last elements of $\lambda$ and non-zeros in other positions. As a result, it lead to zeros at the last elements of $\beta$. In this point of view, in many situations like this, it is not hard to combine a convex constraint on $\hat{\lambda}$, but it may not be available to do the same directly on $\beta$.

In our paper, the sets $\Lambda$ we considered have the expression as follows:

$$\Lambda = \{\lambda \in \mathbb{R}^p_+: A\lambda = S\}$$

where $A$ is a $k \times p$ matrix and $S$ is a convex set. Two main selections of interest are when $S$ is the unit ball of a norm or a convex cone. We refer to the corresponding set $\Lambda$ as norm constraint set or conic constraint, respectively. Two specific examples are considered to respect the sparsity patterns with each selection.

For conic constraint sets, we select $S = \mathbb{R}^k_+$, and thus $\Lambda = \{\lambda \in \mathbb{R}^p_+: A\lambda \geq 0\}$. For the norm constraint sets, we select $S$ as the $l_1$-unit ball and $\Lambda$ is the edge map of an ontology graph $G$ with edge set $E$, i.e., $\Lambda = \{\lambda \in \mathbb{R}^p_+: \sum_{(v_i, v_j) \in E} |\lambda_i - \lambda_j| \leq 1\}$. Furthermore, if $G$ is a one dimension grid ontology graph, then it can be expressed as $\Lambda = \{\lambda \in \mathbb{R}^p_+: \sum_{i=1}^{p-1} |\lambda_i - \lambda_{i+1}| \leq 1\}$.

3.3. Calculating the Proximity Operator

In this subsection, we give the strategies to solve problem (4) using the accelerated first-order technologies. This tricks depend on the proximity operator computation restricted to $\mathbb{R}^p \times \Lambda$.

By virtue of its definition, the proximal operator of $\Gamma$ at $(\alpha, \mu) \in \mathbb{R}^p \times \mathbb{R}^p$ is the solution of the following ontology problem

$$\inf_{(\beta, \lambda)} \frac{1}{2} \|\beta - \mu\|^2 + \rho \Gamma(\beta, \lambda): \beta \in \mathbb{R}^p_+, \lambda \in \Lambda$$

$$\rho \Gamma(\beta, \lambda): \beta \in \mathbb{R}^p_+, \lambda \in \Lambda$$

For given $\lambda$, a calculating gets that the objective function in (6) reaches its minimum at

$$\beta_i(\lambda) = \frac{\alpha \lambda_i}{\lambda_i + \rho}.$$
Specifically, let $\theta_i$ be the parameter defined by $\theta_i = 1$ and 
\[
\frac{1 - \theta_i}{\theta_i^2} = \frac{1}{\theta_i^2},
\]
then the optimal update is 
\[
w_{s+1} \leftarrow u_{s+1} + \theta_i \left( \frac{1}{\theta_i} - 1 \right) (u_{s+1} - u_i).
\]

By computing the proximity operator of \(L^T B\), we rephrased the algorithm using the sequence \(\pi_i = 1 - \theta_i + \sqrt{1 - \theta_i} = 1 - \theta_i + \frac{\theta_i}{\theta_i + 1}\). For each iteration, using Picard-Opial process, \(H_i\) is denoted as 
\[
H_i(x) = (I - \text{prox}_{\frac{L}{\theta_i}})((I - \frac{L}{\theta_i}) BB^T)^* - \frac{1}{L} B(\nabla l(w_i) - Lw_i)).
\]

Also, at stage \(s\) of the proximity calculating is 
\[
v_{s+1} = k v_s + (1 - \kappa) H_i(v_s).
\]
Furthermore, the fixed point process associated with the assignement of \(u\) are equivalent to 
\[
u_{s+1} \leftarrow \text{prox}_{\frac{1}{L}} \frac{1}{L} \nabla l(w_i).\]
Let \(T\) be the total times of iteration. Then the computation complexity of above algorithm is \(O(\frac{1}{T^2})\).

4. SIMULATION STUDIES

In this section, we designed two simulation experiments which are related to ontology similarity measure and ontology mapping, respectively. Note that after getting the sparse vector \(\beta\), the ontology function then is derived by 
\[
f_\beta(v) = \sum_{i=1}^n v_i \beta_i \text{ such that the noise term } \delta \text{ is ignored.}
\]

4.1. Experiment on Plant Data

In this subsection, we use “PO” ontology \(O_1\), constructed in http://www.plantontology.org. (Fig. 1 shows the basic structure of \(O_1\), to check the efficiency of our new algorithm in ontology similarity measuring. Similarly, the \(P@N\) standard [23] is used for this experiment. Moreover, the ontology methods in Wang et al. [11], Huang et al. [12] and Gao and Liang [13] are employed to the “PO” ontology. We calculate the accuracy by the three algorithms, and at last compare the results with the new algorithm. Part of the data can be referred to Table 1.

![Fig. (1). “PO” ontology O1.](image)

### Table 1. The experiment data for ontology similarity measure.

<table>
<thead>
<tr>
<th></th>
<th>(P@3) Average Precision Ratio</th>
<th>(P@5) Average Precision Ratio</th>
<th>(P@10) Average Precision Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Algorithm</td>
<td>49.04%</td>
<td>57.85%</td>
<td>71.01%</td>
</tr>
<tr>
<td>Algorithm in [11]</td>
<td>45.49%</td>
<td>51.17%</td>
<td>58.59%</td>
</tr>
<tr>
<td>Algorithm in [12]</td>
<td>42.82%</td>
<td>48.49%</td>
<td>56.32%</td>
</tr>
<tr>
<td>Algorithm in [13]</td>
<td>48.31%</td>
<td>56.35%</td>
<td>68.71%</td>
</tr>
</tbody>
</table>
Gao and Liang [13]. In particular, such precision ratios are increasing clearly as \( N \) increases. Thus, our algorithm is superior to the method presented by Wang et al., [11], Huang et al., [12] and Gao and Liang [13].

### 4.2. Experiment on University Data

“University” ontologies \( O_2 \) and \( O_3 \), whose structures are presented in Figs. (2 and 3) separately, are used for our last experiment. This experiment tends to determine ontology mapping between \( O_2 \) and \( O_3 \) by means of our ontology algorithm. \( P@N \) criterion is used to measure the equality of the experiment as well. Ontology algorithms in Wang et al., [11], Huang et al., [12] and Gao and Liang [13] are chosen to be applied to university ontologies as comparisons with the results obtained by our algorithm. In the end, through the comparisons of the precision ratio obtained by our algorithm
and the four methods, we get several experiment results which can be referred to Table 2.

The experiment results in Table 2 show that, our algorithm turns out to be more efficient 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.

CONCLUSION

In this paper, the new ontology framework and its optimal technologies are presented for ontology sparse vector computation. The new iterative computation algorithm is based on proximity operator computation, Picard-Opial process and accelerated first-order tricks. At last, simulation data shows that our new algorithm has high efficiency in plant science and university ontologies. The ontology sparse algorithm raised in our paper illustrates the promising application prospects for multiple disciplines.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We thank the reviewers for their constructive comments in improving the quality of this paper. This work was supported in part by the Key Laboratory of Educational Informatization for Nationalities, Ministry of Education, NSFC (No.11401519), and the PhD initial funding of the second author.

REFERENCES