Ontology Vector Learning Scheming Using General Lasso Method and Applied to Multiple Disciplines

Linli Zhu[1], Xiaozhong Min[1], Wei Gao[2] and Haixu Xi[1]

[1] School of Computer Engineering
Jiangsu University of Technology
Changzhou, Jiangsu 213001, China
[2] School of Information
Yunnan Normal University
Kunming, Yunnan 650500, China
zhulinli@jsut.edu.cn[1],gaowei@ynnu.edu.cn[2]

Abstract

In information retrieval and other computer applications, ontology acts as an effective role to retrieve the concepts that have highly semantic similarity with the original query concept, and return the results to the user. Ontology mapping is used to connect the relationship between different ontologies, and similarity computation is the essence of such applications. In this article, we present a new ontology sparse vector scheming for ontology similarity measure and ontology mapping in terms of general fusion lasso. The solution of ontology optimization problem is obtained via learning its Lagrangian version. The implementation procedure is based on gradient computating and fusion step, and the parameters in the ontology framework are chosen by means of cross-vaildation and bayesian information criterion. The simulation experiment results show that the newly proposed method has high efficiency and accuracy in ontology similarity measure and ontology mapping in multiple disciplines.

Keywords: Ontology, Similarity measure, Ontology mapping, General lasso, Cross-vaildation

1. Introduction

Ontology, known as a kind of knowledge representation and conceptual shared model is used in image retrieval, knowledge management and information retrieval search extension. With high effectiveness as a concept semantic model, ontology is also extensively used in other research fields such as social science, medical science, biology science, pharmacology science and geography science (for instance, see [1-5]).

The ontology model is in fact a graph $G=(V,E)$, in which the vertex v represents a concept and each edge $e=v_i v_j$ represents a relationship between concepts v_i and v_j. To find a similarity function *Sim*: $V \times V \rightarrow \Box^+ \cup \{0\}$ is the target of ontology similarity measure as it maps each pair of vertices to a real number. Bridging the link between two or more ontologies is the purpose of ontology mapping. İt is necessary to make G_1 and G_2 be two ontology graphs corresponding respectively to ontology O_1 and O_2. There is a need to find a set $S_v \subseteq V(G_2)$ for each $v \in G_1$ as concepts corresponding to vertices in S_v are semantically close to the concept corresponding to v. In order to get such mapping, compute the similarity $S(v,v_j)$ where $v_j \in V(G_2)$ and choose a parameter $0 < M < 1$ for each $v \in G_1$. S_v is a collection like the element in S_v satisfies $S(v,v_j) \geq M$. İt can be seen from this point of view that the essence of ontology mapping is to obtain a similarity function S and to select a suitable parameter M.

There are several effective learning tricks in ontology similarity measure and ontology mapping. [6] proposed that the function of a score which maps 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. [7] presented ontology similarity can be calculated by in fast ontology algorithm in a short time. [8] raised that optimizing NDCG measure determines the optimal ontology function and this idea can be applied in education of physics. [9] used the regression approach and deduced the ontology function. İn the research of [10], new ontology mapping algorithm with the use of harmonic analysis and diffusion regularization on hypergraph were presented. New ontology similarity computation technology such as the new calculation model was proposed in the research of [11], which put the operational cost of real implement into consideration. The ontology similarity measuring and ontology mapping algorithms on the basis of minimum error entropy criterion were presented in the latest research of [12]. Several theoretical analysis of ontology algorithm by [13-16] can be referred to.

New ontology similarity computation and ontology mapping algorithms based on general fused lasso tricks are determined in this paper. Furthermore, we design four experiments to show the efficiency of the algorithm in the applications.

2. Setting

Let V be instance space. A p dimension vector expresses the information of each vertex in ontology graph (its name, instance, attribute, structure, and semantic information of the concept which corresponds to the vertex and is contained in name and attribute components of its vector are included). Let $v = \{v_1, \cdots, v_p\}$ be a vector corresponding to a vertex v. To facilitate the representation, the notations are slightly confused by us and v is used to denote both the ontology vertex and its corresponding vector. The aim of ontology learning algorithms is to get an optimal ontology (score) function $f: V \to \Box$, the difference between two real numbers that they correspond to judges the similarity between two vertices. Dimensionality reduction is the essence of this algorithm, *i.e.*, p dimension vector is expressed by one dimension vector. An ontology function f is specifically a dimensionality reduction function $f: \Box^p \to \Box$.

As all the information of the vertex concept, attribute and the neighborhood structure in the ontology graph are contained within the vector corresponding to a vertex of ontology graph, high dimension is always with it. In the biological ontology, for example, the information of all genes is contained in a vector. In addition, ontology structure can be very complicated because of ontology graph with large number of vertices, and the GIS (Geographic Information System) ontology can be the most typical example. The fact that the similarity calculation of ontology application is very large may be led to by these factors. But in fact, only a small part of the vector components determines the similarity between the vertices. In the application of biological ontology, for example, a small number of genes often causes a genetic disease while most of the other genes are irrelevant. Moreover, in the geographic information system ontology, if an accident happens in a place and casualties are caused, it is necessary for us to find the nearest hospital regardless of schools and shops nearby, *i.e.*, neighborhood information that meets specific requirements on the ontology graph needs to be found by us. Therefore, research the sparse ontology algorithm attracts tremendous academic and industrial interests.

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

$$f_w(v) = \sum_{i=1}^{p} v_i w_i + \delta \quad . \tag{1}$$

Here $w = (w_1, \cdots, w_p)$ is a sparse vector and δ is a noise term. The sparse vector w is to shrink irrelevant component to zero. To determine the ontology function f, sparse vector w should be learnt at first.

One popular model with the penalize term via the l_1-norm of the unknown sparse vector $w \in \square^p$:

$$Y_w = l(w) + \lambda \|w\|_1 \quad \lambda > 0 . \tag{2}$$

Here, $\lambda > 0$ is a regularization parameter (or, balance parameter) and l is the principal function (in many articles, it is called loss function) to measure the quality of w. The balance term $\lambda \|w\|_1$ measures the sparsity of sparse vector w.

3. Algorithm Description

In what follows, we assume that we observe ontology data $(v_1, y_1), \ldots, (v_m, y_m)$, where $v_j = (v_{j1}, \cdots, v_{jp})^{\mathrm{T}} \in \square^p$ and y_j are the response variables. Then, the ontology linear model corresponding to ontology function expression can be denoted as (for $j=1,2,\ldots,m$) $y_j = \sum_{i=1}^p w_i v_{ij} + \delta_j$, satisfies $E(\delta_j) = 0$ and $\mathrm{Var}(\delta_j) = \sigma^2$. Furthermore, we suppose that the response variable is centered and the predictors are standardized, i.e., $\sum_{j=1}^m y_j = 0$, $\sum_{j=1}^m v_{ij} = 0$ and $\sum_{j=1}^m v_{ij}^2 = 1$ are established for each $i=1,\ldots,p$.

The ontology sparse vector learning framework in our setting is obtained by minimizing the following euclidean norm

$$\hat{w} = \underset{w}{\arg\min} \left\| y - \sum_{i=1}^p w_i v_i \right\|_2^2 , \text{s.t.} \sum_{i=1}^p |w_i| \le r,$$

where $y = (y_1, \cdots, y_m)$, $v_j = (v_{1i}, \cdots, v_{mi})$ and $r \in \square^+$ is a parameter to control the sparse of ontology vector w (it restricts the number of non-zero components in w).

The view of elastic net (see [17-21] for more details), ontology sparse vector learning problem can be denoted as the following constrained least squares optimization problem:

$$\hat{w} = \underset{w}{\arg\min} \left\| y - \sum_{i=1}^p w_i v_i \right\|_2^2 , \text{ s.t. } \alpha \sum_{i=1}^p |w_i| + (1-\alpha) \sum_{i=1}^p w_i^2 \le r.$$

The first part causes a sparse ontology solution of these averaged predictors while the second part impels highly correlated predictors to be averaged.

The elastic-net based scheme can be revised as the following clustering shrinkage algorithm:

$$\hat{w} = \underset{w}{\arg\min} \left\| y - \sum_{i=1}^p w_i v_i \right\|_2^2 , \text{ s.t. } \sum_{i=1}^p |w_i| + c \sum_{i<k}^p \max\{|w_i|, |w_k|\} \le r.$$

The fused lasso based ontology optimization algorithm is determined by solving (refer to [22-26] see more on fused lasso technology and its applications)

$$\hat{w} = \underset{w}{\arg\min} \left\| y - \sum_{i=1}^p w_i v_i \right\|_2^2 , \text{ s.t. } \sum_{i=1}^p |w_i| \le r_1 \text{ and } \sum_{i=2}^p |w_i - w_{i-1}| \le r_2 .$$

The clustered lasso as an extension of the fused lasso was introduced by [27]:

$$\hat{w} = \underset{w}{\arg\min} \left\| y - \sum_{i=1}^{p} w_i v_i \right\|_2^2, \text{ s.t. } \sum_{i=1}^{p} |w_i| \le r_1 \text{ and } \sum_{i<k} |w_i - w_k| \le r_2.$$

[28] encapsulated the former tricks in the framework of the general lasso. The corresponding ontology problem can be expressed as:

$$\hat{w} = \underset{w}{\arg\min} \left\| y - \sum_{i=1}^{p} w_i v_i \right\|_2^2, \text{ s.t. } \|\boldsymbol{Dw}\|_1 \le r,$$

where $\boldsymbol{D} \in \Box^{m \times p}$ is a particular balance matrix. For example,

$$\boldsymbol{D} = \begin{pmatrix} \mathrm{I}_n \\ \lambda \boldsymbol{F} \end{pmatrix}, \boldsymbol{F} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$

In this paper, our ontology framework is followed by the general fused lasso technologies, but new calculating tricks are used in our ontology scheme. Let d be a thresholding parameter and $\frac{1}{d} \le \alpha \le 1$. We use least-squares and linear combination balance term in our ontology optimization framework:

$$\hat{w} = \underset{w}{\arg\min} \left\| y - \sum_{i=1}^{p} w_i v_i \right\|_2^2, \tag{3}$$

$$\text{s.t. } \alpha \sum_{i=1}^{p} |w_i| + (1-\alpha)\sum_{i<k} |w_i - w_k| \le r.$$

The balance term in the model above implies a geometric explanation of the constraint field which is unable to manifest the expression of the balance term. We emphasize that ontology problem (3) can be re-written as the following Lagrangian version:

$$\underset{w}{\arg\min}\left\{ \left\| y - \sum_{i=1}^{p} w_i v_i \right\|_2^2 + \lambda\left(\alpha \sum_{i=1}^{p} |w_i| + (1-\alpha)\sum_{i<k} |w_i - w_k|\right) \right\}. \tag{4}$$

Now, we discuss the solution and persistence property of ontology problem (3) and (4).

Let $R(\boldsymbol{w}) = \frac{1}{m} E \left\| \boldsymbol{y} - \sum_{i=1}^{p} w_i v_i \right\|_2^2$, $\hat{R}(\boldsymbol{w}) = \frac{1}{m} \left\| \boldsymbol{y} - \sum_{i=1}^{p} w_i v_i \right\|_2^2$, $\boldsymbol{w}^* = \underset{w \in \Theta_m}{\arg\min R(\boldsymbol{w})}$, where Θ_m is a constraint region of $\boldsymbol{w}$. Then, $\hat{w}$ is persistent if $\hat{R}(\hat{w}) - R(\boldsymbol{w}^*) \xrightarrow{p} 0$.

In view of the technologies raised in [29], set $z = (Z_0, \ldots, Z_p) = (Y, V_1, \ldots, V_p)$. Set $w_0 = -1$. Then, we infer $R(\boldsymbol{w}) = \sum_{i=0}^{p}\sum_{k=0}^{p} w_i w_k E(Z_i Z_k)$, $\hat{R}(\boldsymbol{w}) = \frac{1}{m}\sum_{j=1}^{m}\sum_{i=0}^{p}\sum_{k=0}^{p} w_i w_k Z_{ij} Z_{jk}$, and

$$\left| \hat{R}(\boldsymbol{w}) - R(\boldsymbol{w}^*) \right| \le \max_{i,k} \left| \frac{1}{m}\sum_{j} Z_{ij} Z_{jk} - E(Z_i Z_k) \right| \sum_{i}\sum_{k} |w_i w_k|$$

$$= \|\boldsymbol{w}\|_2^2 \max_{i,k} \left| \frac{1}{m}\sum_{j} Z_{ij} Z_{jk} - E(Z_i Z_k) \right|.$$

Suppose that

$$\max_{i,k}\left|\frac{1}{m}\sum_{j=1}^{m}Z_{ij}Z_{jk}-E(Z_iZ_k)\right|\underset{=}{}O(\sqrt{\frac{\log m}{m}})$$. Hence, we deduce that

$$\sup_{w\in\Theta_m^0}\left|\hat{R}(w)-R(w^*)\right|\underset{\le}{}\sup_{w\in\Theta_m}\|w\|_1^2\, O(\sqrt{\frac{\log m}{m}})$$. Let $\Theta_m^0=\{w:\sum_{i=1}^{p}|w_i|\le r_m\}$. Then, the

sufficient condition for persistence of our ontology problem is

$$\sup_{w\in\Theta_m^0}\|w\|_1^2\underset{=}{}r_m^2\underset{=}{}O(\frac{m}{\log m})^{\frac{1}{2}}$$, Set constraint regions of general fussed lasso based ontology

problem as follows:

$$\Theta_m=\{v^{\mathrm{T}}w:\alpha\sum_{i=1}^{p}|w_i|+(1-\alpha)\sum_{i<k}|w_i-w_k|\le r_m,\frac{1}{\sqrt{p}}\le\alpha\le 1\}$$.

$$\Theta_m'=\{v^{\mathrm{T}}w:\alpha\sum_{i=1}^{p}|w_i|+(1-\alpha)\sum_{i<k}\max\{|w_i|,|w_k|\}\le r_m,0\le\alpha\le 1\}$$.

$$\Theta_m''=\{v^{\mathrm{T}}w:\alpha\sum_{i=1}^{p}|w_i|+(1-\alpha)\sum_{i=1}^{p}w_i^2\le r_m\}$$.

We can verifies that $\|w\|_1^2$ reaches the maximum value if $w_1=w_2=\ldots=w_p$. From this

point of view, we have $\sup_{w\in\Theta_m'}\|w\|_1^2=(pr_m)^2$ and $\sup_{w\in\Theta_m''}\|w\|_1^2=pr_m^2$. Thus, the general fused

lasso based ontology problem is persistent if $r_m=O(\frac{m}{p^2\log m})^{\frac{1}{2}}$. The maximum value is

yielded if $w_1=w_2=\ldots=w_p=\frac{r_m}{p\alpha}$. And hence, $\sup_{w\in M_n}\|w\|_2^2=(\frac{r_m}{\alpha})^2=pr_m^2$ if α take the

value $\frac{1}{\sqrt{p}}$.

Next, we introduce the trick to implement the ontology problem by minimizing the prediction error with cross-validation. Since solving the following ontology procedure,

$$\hat{w}=\underset{w}{\arg\min}\left\|y-\sum_{i=1}^{p}w_iv_i\right\|_2^2$$, s.t. $\alpha\sum_{i=1}^{p}|w_i|+(1-\alpha)\sum_{i<k}|w_i-w_k|\le r$, is equivalent to

determine its corresponding Lagrangian counterpart

$$g(w)=\frac{1}{2}\left\|y-\sum_{i=1}^{p}w_iv_i\right\|_2^2+\lambda_1\sum_{i=1}^{p}|w_i|+\lambda_2\sum_{i<k}|w_i-w_k| \quad , \quad (5)$$

where $\lambda_1=\lambda\alpha$, $\lambda_2=\lambda(1-\alpha)$ and $\lambda>0$.

We should take a descent step to solve (5). Let $\tilde{w}_i$ be the current estimates of the w_i and sgn(v) be the subgradient of $|v|$. Then, the derivative of Lagrangian version (5) with respect to w_k given $w_i=\tilde{w}_i$, $i\ne k$ can be expressed as

$$\frac{\partial g(w)}{\partial w_k}=v_k^{\mathrm{T}}v_kw_k-(y-\sum_{i\ne k}\tilde{w}_iv_i)^{\mathrm{T}}v_k+\lambda_1\operatorname{sgn}(w_k)$$

$$+\lambda_2 \sum_{i=1}^{k-1} \text{sgn}(\tilde{w}_i - w_k) + \lambda_2 \sum_{i=k+1}^{p} \text{sgn}(w_k - \tilde{w}_i)$$

, (x) Clearly, the derivative version is piecewise linear in w_k with breaks at $\{0, \tilde{w}_i, i \neq k\}$, or wise $w_k \notin \{0, \tilde{w}_i, i \neq k\}$.

If there exists a solution for $\dfrac{\partial g(\boldsymbol{w})}{\partial w_k} = 0$, an interval (c_1, c_2) can be determined as which contains its solution, and we further have

$$\tilde{w}_k = \text{sgn}\{\tilde{y}^T v_k - \lambda_2(\sum_{i<k} s_{ik} + \sum_{i>k} s_{ki})\} \times \frac{(\left|\tilde{y}^T v_k - \lambda_2(\sum_{i<k} s_{ik} + \sum_{i>k} s_{ki})\right| - \lambda_1)_+}{v_k^T v_k},$$

where $s_{ik} = \text{sgn}(\tilde{w}_i - \dfrac{c_1 + c_2}{2})$ and $\tilde{y} = y - \sum_{i \neq k} \tilde{w}_i v_i$.

If we can't find the solution for $\dfrac{\partial g(\boldsymbol{w})}{\partial w_k} = 0$, then we set

$$\tilde{w}_k = \begin{cases} \tilde{w}_l & \text{if } g(\tilde{w}_l) = \min\{g(0), g(\tilde{w}_i), \text{ for } i \neq k\} \\ 0 & \text{if } g(0) \leq g(\tilde{w}_i) \text{ for each } i \neq k \end{cases}.$$

For the situation that the descent step lost to reduce the function $g(\boldsymbol{w})$, we should discuss the fusion of pairs of w_k. For every single pair (k,l), $l \neq k$, we consider the constraint $w_k = w_l = \gamma$ and discuss a descent change in γ. By setting $\tilde{y} = y - \sum_{i \neq k,l} \tilde{w}_i v_i$, the derivative of (5) with respect to γ then becomes

$$\frac{\partial g(\boldsymbol{w})}{\partial \gamma} = (v_k^T v_k + v_l^T v_l)\gamma - \tilde{y}^T(v_k + v_l) + 2\lambda_1 \text{sgn}(\gamma)$$

$$+ 2\lambda_2 \sum_{i<k,l} \text{sgn}(\tilde{w}_i - \gamma) + 2\lambda_2 \sum_{i>k,l} \text{sgn}(\gamma - \tilde{w}_i).$$

The change $w_k = w_l = \gamma$ is permitted if the optimal value of γ deduced by virtue of the descent step decreases the objective function.

The last part in this section is to discuss the selection of parameters α and γ. Our technology is based on K-fold cross-validation which divides randomly the ontology data into K roughly with equal size and disjoint subcollections D_k ($k=1, \ldots, K$) satisfies $\bigcup_{k=1}^{K} D_k = \{1,2,\ldots, n\}$ (readers can refer to [30-35] for more details about the method of cross-validation). Let $\hat{w}_i^{(-k)}(\alpha, \lambda)$ be the estimate of w_i for a fixed parameter α and γ in terms of the ontology data set without D_k. Then, the statistic error for cross-validation is defined by $CV(\alpha, \lambda) = \sum_{k=1}^{K} \sum_{j \in D_k} (y_j - \sum_{i=1}^{p} \hat{w}_i^{(-k)}(\alpha, \lambda)v_{ij})^2$.

Generalized cross-validation and Bayesian information criterion (BIC) (see [36-40] for more details) are other popular tricks. Let $\hat{w}_i(\alpha, \lambda)$ be the estimate of w_i for a fixed parameter α and γ, dg be the degree of freedom which is a measure of model

complexity, and $RSS(\alpha,\lambda) = \sum_{j=1}^{m}(y_j - \sum_{i=1}^{p}\hat{w}_i(\alpha,\lambda)v_{ij})^2$. Then, generalized cross-validation and Bayesian information criterion can be expressed as follows

$$GCV(\alpha,\lambda) = \frac{RSS(\alpha,\lambda)}{m-dg} , BIC(\alpha,\lambda) = m\log(RSS(\alpha,\lambda)) + \log m \times dg .$$

4. Experiments

In this section, we design four simulation experiments relevance ontology similarity measure and ontology mapping below. In order to adjacent to the setting of ontology algorithm, a vector with d dimension to express each vertex's information is used. The information such as name, instance, attribute and structure of vertex is contained in this vector. Here the set of its reachable vertex in the directed ontology graph is represented by the instance of vertex.

In the following four experiment, the effectiveness of main ontology algorithm in is

verified in our paper. After getting the sparse vector w , then the $f_w(v) = \sum_{i=1}^{p}v_i w_i$. causes the ontology function.

4.1. Experiment on Biology Data

"Go" ontology O_1 which was constructed in http: //www. geneontology. org. (Figure 1 shows the basic structure of O_1) is used for our experiment. We use $P@N$ (Precision Ratio, see [41] for more detail) to measure the equality of the experiment. First of all, the closest N concepts for every vertex was given on the ontology graph by expert, and then the first N concepts for every vertex on ontology graph by the algorithm and compute the precision ratio was obtained by us. [7-8] and [13] employed ontology algorithms to "Go" ontology, and the precision ratio which we got from four methods was compared. Several experiment results can be referred to Table 1.
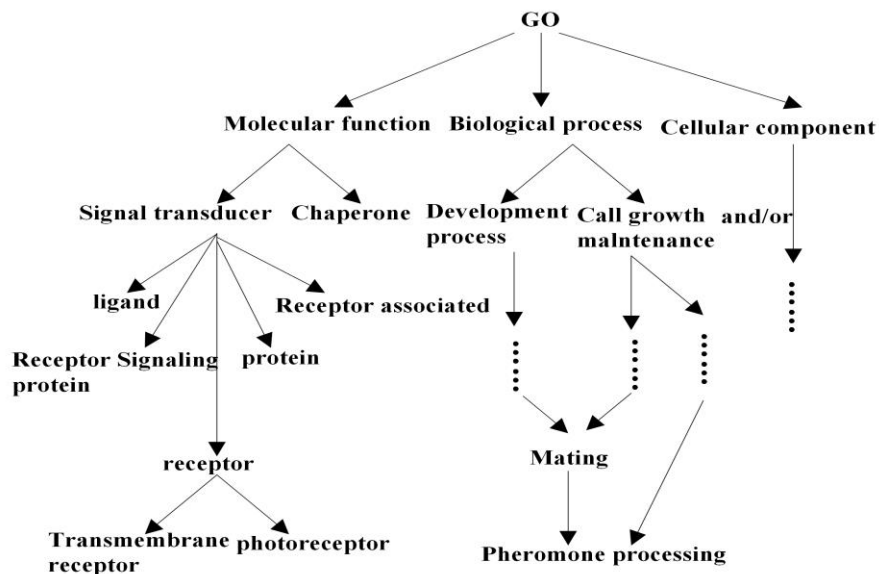

Figure. 1. "Go" Ontology

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
Our Algorithm	47.24%	54.62%	64.95%	79.16%
Algorithm in[7]	46.38%	53.48%	62.34%	74.59%
Algorithm in[8]	43.56%	49.38%	56.47%	71.94%
Algorithm in [13]	42.13%	51.83%	60.19%	72.39%

When N= 3, 5, 10 or 20, the precision ratio by virtue of our algorithm is higher than the precision ratio determined by algorithms as [7-8, 13] proposed in their research. Particularly, when N increases, such precision ratios will apparently increase. We can concluded that the algorithm described in our paper is superior to the method proposed by [7-8, 13].

4.2. Experiment on Physical Education Data

Physical education ontologies O_2 and O_3 (the structures of O_2 and O_3 are presented in Figure 2 and Figure 3 respectively) were used for our second experiment. Determining the ontology mapping between O_2 and O_3 We applied P@N criterion to measure the equality of the experiment. The closest N concepts for each vertex on the ontology graph with the help of experts was given by us, and then the first N concepts for every vertex on ontology graph by the algorithm and compute the precision ratio was obtained. [7-8, 10] also employed ontology algorithms to "physical education" ontology, and we made a comparison among the precision ratios which we get from four methods. Several experiment results can be referred to Table 2.
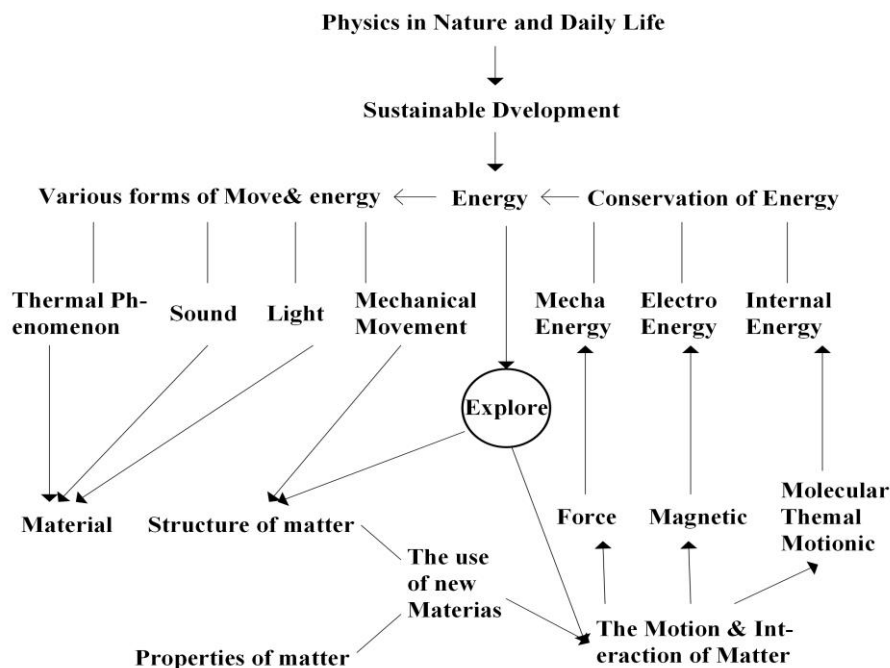

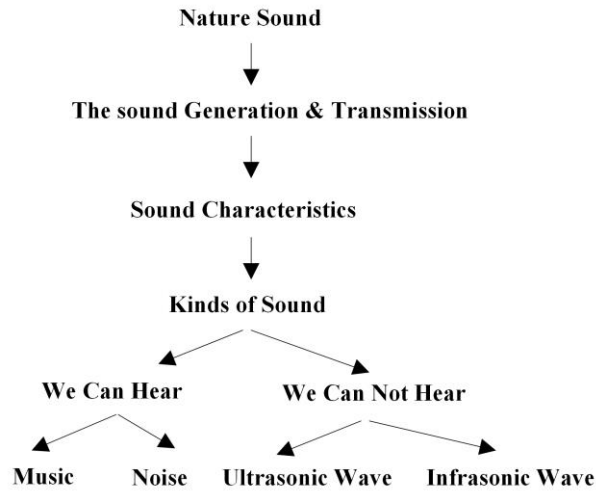**Figure 2. "Physical Education" Ontology O₂**

Figure 3. "Physical Education" Ontology O_3

Table 2. The Experiment Data for Ontology Mapping

	P@1 average precision ratio	*P@3* average precision ratio	*P@5* average precision ratio
Our Algorithm	69.13%	78.49%	93.55%
Algorithm in [7]	61.29%	73.12%	79.35%
Algorithm in [8]	69.13%	75.56%	84.52%
Algorithm in [10]	67.74%	77.42%	89.68%

That our algorithm is more efficiently than algorithms raised in [7-8, 10] especially when N is sufficiently large is revealed by the experiment results in Table 2.

4.3. Experiment on Plant Data

In this subsection, we use "PO" ontology O_4 which was constructed in http://www.plantontology.org. (Figure 4 shows the basic structure of O_4) to test the efficiency of our new algorithm for ontology similarity measuring. We use the *P@N* standard again for this experiment. Furthermore, ontology method in [6-8] to the "PO" ontology is applied in our experiment. The accuracy by these three algorithms is calculated and the result to algorithm rose in our paper is compared, part of the data can be referred to Table 3.
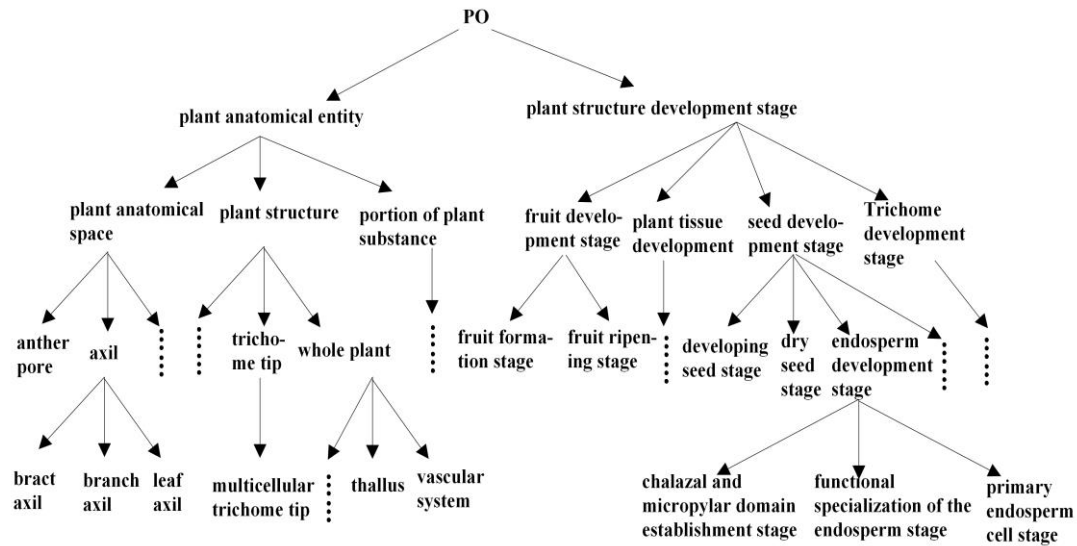
Figure 4. "PO" ontology O_4

Table 3. The Experiment Data for Ontology Similairty Measure

	$P@3$ average precision ratio	$P@5$ average precision ratio	$P@10$ average precision ratio
Our Algorithm	48.37%	57.13%	71.12%
Algorithm in [6]	45.49%	51.17%	58.59%
Algorithm in [7]	42.82%	48.49%	56.32%
Algorithm in [8]	48.31%	56.35%	68.71%

When N= 3, 5, or 10, the precision ratio in terms of our algorithm is higher than the precision ratio determined by algorithms that [6-8] proposed in their research. In particular, such precision ratios are increasing apparently as N increases. İt can be concluded that the algorithm described in our paper is superior to the method that [6-8] proposed in their research.

4.4. Experiment on Humanoid Robotics Data

Humanoid robotics ontologies O_5 and O_6 (constructed by [42], and the structures of O_5 and O_6 are presented in Figure 5 and Figure 6 respectively) were used for our last experiment. This experiment is to determine ontology mapping between O_5 and O_6 via similarity matrix which Algorithm 1 deduces. We use $P@N$ criterion to measure the equality of the experiment. We employ ontology algorithms in [43, 8, 10] to humanoid robotics ontologies, and the precision ratio which we get from four methods are compared. Several experiment results can be referred to Table 4.

The experiment results in Table 4 reveal that our algorithm works with more efficiency than algorithms [43, 8, 10] raised in their research especially when N is sufficiently large.

Table 4. The Experiment Data for Ontology Mapping

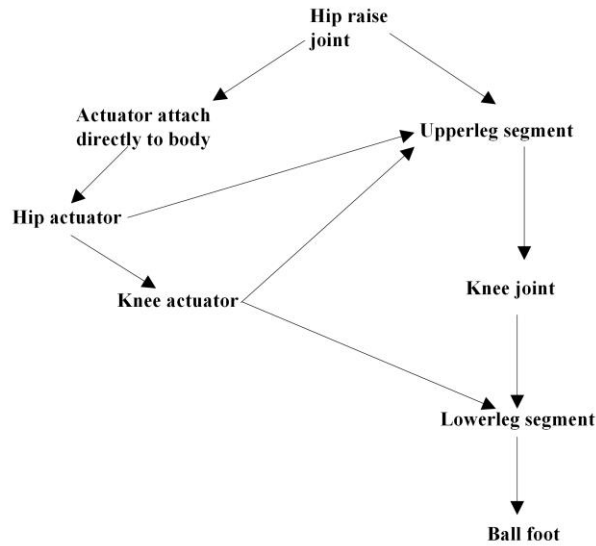|	$P@1$ average precision ratio	$P@3$ average precision ratio	$P@5$ average precision ratio
Our Algorithm	27.78%	53.70%	68.89%
Algorithm in [43]	27.78%	48.15%	54.44%
Algorithm in [8]	22.22%	40.74%	48.89%
Algorithm in [10]	27.78%	46.30%	53.33%

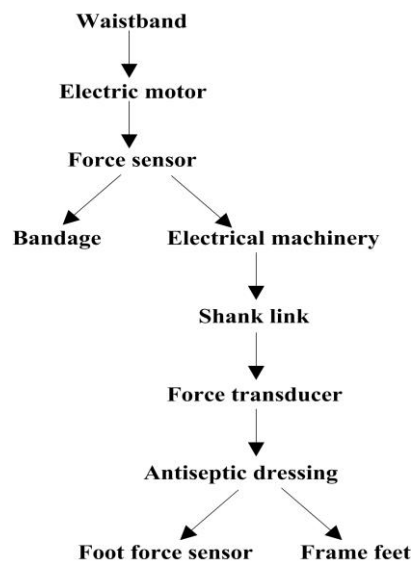**Figure 5. "Humanoid Robotics" Robotics" Ontology O_5**

Figure 6. "Humanoid Ontology O_6.

5. Conclusion

In our article, a fused lasso based computation technology is manifested for ontology similarity measure and ontology mapping application. The detailed steps rely on the Lagrangian alternation, derivative calculation and parameter selection. At last, simulation

data shows that our ontology scheming has high efficiency in biology, physics education, plant science and humanoid robotics. The ontology sparse algorithm raised in our paper illustrates the promising application prospects in multiple disciplines.

Acknowledgment

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, **(2013)**, pp. 3211-3219.

[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, **(2014)**, pp. 966- 974.

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

[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, **(2014)**, pp. 1598-1628.

[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, **(2014)**. pp. 1061-1083.

[6] 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, **(2010)**; Wuhan, China.

[7] 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, **(2011)**, pp. 54-59.

[8] 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, **(2011)**, pp. 415-421.

[9] 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, **(2012)**, pp. 107-112.

[10] 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, **(2013)**, pp. 592-598.

[11] 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, **(2013)**, pp. 1572-1577.

[12] 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, **(2014)**, pp. 3793-3806.

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

[14] W. Gao and T. W. Xu, "Characteristics of optimal function for ontology similarity measure via multi-dividing", Journal of networks, vol. 7, no. 8, **(2012)**, pp. 1251-1259.

[15] L. Yan, W. Gao and J. Li, "Piecewise function approximation and vertex partitioning schemes for multi-dividing ontology algorithm in AUC criterion setting (II)", Journal of Applied Science, vol. 13, no. 16, **(2013)**, pp. 3257-3262.

[16] 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, **(2014)**, pp. 226-231.

[17] B. J. Marafino, W. J. Boscardin and R. A. Dudley, "Efficient and sparse feature selection for biomedical text classification via the elastic net: application to ICU risk stratification from nursing notes", Journal of Biomedical Informatics, vol. 54, **(2015)**, pp. 114-120.

[18] N. Vlassis and E. Glaab, "GenePEN: analysis of network activity alterations in complex diseases via the pairwise elastic-net", Statistical Applications in Genetics and Molecular Biology, vol. 14, no. 2, **(2015)**, pp. 221-224.

[19] S. Ito, T. Kinoshia and W. G. Bao, "Hydrodynamic behaviors of an elastic net structure", Ocean Engineering, vol. 92, **(2014)**, pp. 188-197.

[20] H. F. Schepker, C. Bockelmann, A. Dekorsy, A. Bartels, D. Trede and K. S. Kazimierski, "C-curve: a finite alphabet based parameter choice rule for elastic-net in sporadic communication", IEEE Communications Letters, vol. 18, no. 8, **(2014)**, pp. 1443-1446.

[21] D. M. Percival, D. B. Percival, D. W. Denbo, E. Gica, P. Y. Huang, H. O. Mofjeld and M. C. Spillane, "Automated tsunami source modeling using the sweeping window positive elastic-net", Journal of the American Statistical Association, vol. 109, no. 506, **(2014)**, pp.109:491-499.

[22] D. Yu, J. H. Won, T. Lee, J. Lim and S. Yoon, "High-dimensional fused lasso regression using majorization-minimization and parallel processing", Journal of Computational and Graphical Statistics, vol. 24, no. 1, **(2015)**, pp. 121-153.

[23] N. Chaturvedi, R. X. de Menezes and J. J. Goeman, "Fused lasso algorithm for Cox proportional hazards and binomial logit models with application to copy number profiles", Biometrical Journal, vol. 56, no. 3, **(2014)**, pp. 477-492.

[24] E. R. Lee, H. Noh and B. U. Park, "Model selection via bayesian information criterion for quantile regression models", Journal of the American Statistical Association, vol. 109, no. 505, **(2014)**. pp. 216-229.

[25] F. Heinzl and G. Tutz, "Clustering in linear-mixed models with a group fused lasso penalty", Biometrical Journal, vol. 56, no. 1, **(2014)**, pp. 44-68.

[26] J. D. B. Nelson, "Fused Lasso and rotation invariant autoregressive models for texture classification", Pattern Recognition Letters, vol. 34, no. 16, **(2013)**, pp. 2166-2172.

[27] Y. She, "Sparse regression with exact clustering", Electronic Journal of Statistics, vol. 4, **(2010)**, pp. 1055-1096.

[28] R. J. Tibshirani and J. Taylor, "The solution path of generalized lasso", Annals of Statistics, vol. 39, no. 3, **(2011)**, pp. 1335-1371.

[29] E. Greenshtein and Y. Ritov, "Persistence in high-dimensional linear predictor selection and the virtue of overparametrization", Bernoulli, vol. 10, no. 6, **(2004)**, pp. 971-988.

[30] 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, **(2013)**. pp. 77-85.

[31] 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, **(2013)**, pp. 56-63.

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

[33] 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, **(2013),** pp. 22-31.

[34] 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, no. 7-8, **(2014)**, pp. 1761-1770.

[35] 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, **(2014)**, pp. 64-71.

[36] S. Kawano, "Selection of tuning parameters in bridge regression models via Bayesian information criterion", Statistical Papers, vol. 55, no. 4, **(2014)**, pp. 1207-1223.

[37] C. Cobos, H. Munoz-Collazos, R. Urbano-Munoz, M. Mendoza, E. Leon and E. Herrera-Viedma, "Clustering of web search results based on the cuckoo search algorithm and balanced bayesian information criterion", Information Sciences, vol. 281, **(2014)**, pp. 248-264.

[38] S. H. Lee, D. Yu, A. H. Bachman, J. Lim and B. A. Ardekani, "Application of fused lasso logistic regression to the study of corpus callosum thickness in early Alzheimer's disease", Journal of Neuroscience Methods, vol. 221, **(2014)**, pp. 78-84.

[39] Y. D. Ko, P. Moon, C. E. Kim, M. H. Ham, M. K. Jeong, A. Garcia-Diaz, J. M. Myoung and I. Yun, "Predictive modeling and analysis of HfO2 thin film process based on Bayesian information criterion using PCA-based neural networks", Surface and Interface Analysis, vol. 45, no. 9, **(2013)**, pp. 1334-1339.

[40] S. Watanabe, "A widely applicable bayesian information criterion", Journal of Machine Learning Research, vol. 14, **(2013)**, pp. 867-897.

[41] 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)**.

[42] W. Gao and L. L. Zhu, "Gradient learning algorithms for ontology computing", Computational Intelligence and Neuroscience, http://dx.doi.org/10.1155/2014/438291, **(2014)**.

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