2013

# Ontology Algorithm Based on Semi-supervised Eigenmap Learning

Bo Peng
*School of Information Science and Technology, Yunnan Normal University, Kumming 650500, China*, gaowei@ynnu.edu.cn

Tian-wei Xu
*School of Information Science and Technology, Yunnan Normal University, Kumming 650500, China*, gaowei@ynnu.edu.cn

Zhen Li
*School of Information Science and Technology, Yunnan Normal University, Kumming 650500, China*, gaowei@ynnu.edu.cn

Wei Gao
*School of Information Science and Technology, Yunnan Normal University, Kumming 650500, China*, gaowei@ynnu.edu.cn

# Ontology Algorithm Based on Semi-supervised Eigenmap Learning

*Bo Peng, Tian-wei Xu, Zhen Li and Wei Gao* *

School of Information Science and Technology, Yunnan Normal University, Kumming 650500, China

**Abstract:** Ontology similarity calculation is an important research topic in information retrieval and it is also widely used in computer science. In this paperwe propose new algorithms for ontology similarity measurement and ontology mapping using Semi-supervised Eigenmap Learning method. Via the learning algorithmthe ontology graph is mapped into a line consists of real numbers. The similarity between two concepts then can be measured by comparing the difference between their corresponding real numbers. Two experimental results show that the proposed new algorithm has high accuracy and efficiency on ontology similarity calculation and ontology mapping.

**Keywords:** ontology, ontology mapping, semi-supervised learning, spectral clustering, graph Laplacian, similarity function

## 1 Clustering and Graph Laplacian

Clustering is an important topic in statistics, computer science and many other fields. *Spectral clustering* is a useful family of clustering algorithms which have been applied to some areas involving data analysis and processing. These algorithms are based on graph Laplacians which is a classical topic in spectral graph theory [2] and is introduced to learning theory in [1]. Their goal is to cluster data points according to the values on the data points of a function which is taken to be an eigenfunction of the graph Laplacian with the second smallest eigenvalue.

Let $\mathscr{X}$ be a compact metric space which contains data points, and $P$ be a probability measure on $\mathscr{X}$. A *similarity function* $k : \mathscr{X} \times \mathscr{X} \to \mathbb{R}_+$ is symmetric, continuous and bounded away from 0 by a positive constant $l$. It is used to measure similarity between different data points. Let $\mathbf{x} = \{x_i\}_{i=1}^n$ be a random sample drawn independent according to the probability measure $P$. Then the data similarity matrix is defined by $K_{\mathbf{x}} = [k(x_i, x_j)]_{i,j=1}^n$, and the degree matrix $D_{\mathbf{x}}$ is defined as the $n \times n$ diagonal matrix with diagonal entries $(D_{\mathbf{x}})_{i,i} = \sum_{j=1}^n k(x_i, x_j)$. The (normalized) *graph Laplacian* is defined as the $n \times n$ matrix

$$L_{\mathbf{x}} = I - D_{\mathbf{x}}^{-1/2} K_{\mathbf{x}} D_{\mathbf{x}}^{-1/2}. \qquad (1)$$

Observe that $L_{\mathbf{x}}$ has a trivial smallest eigenvalue 0 with eigenvector $D_{\mathbf{x}}^{\frac{1}{2}}[1, \ldots, 1]^T$ and it is positive semi-definite [1]. The "second smallest eigenvalue" refers to the smallest eigenvalue beyond the trivial eigenvalue 0, and a corresponding eigenvector may be used for clustering data. To see this, we need to introduce some linear operators on $C(\mathscr{X})$, the space of continuous functions on $\mathscr{X}$ with norm $\|\cdot\|_\infty$. Denote $P_n := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$.

Let $p_{\mathbf{x}} := \int_{\mathscr{X}} k(\cdot, y) dP_n(y) = \frac{1}{n} \sum_{i=1}^n k(\cdot, x_i) \in C(\mathscr{X})$ and $p := \int_{\mathscr{X}} k(\cdot, y) dP(y) \in C(\mathscr{X})$ be degree functions. Then the integral operator $T : C(\mathscr{X}) \to C(\mathscr{X})$ associated with $(k, P)$ is defined by

$$T f(x) = \int_{\mathscr{X}} \frac{k(x, y)}{\sqrt{p(x)}\sqrt{p(y)}} f(y) dP(y), \qquad (2)$$

where $x \in \mathscr{X}$, $f \in C(\mathscr{X})$. The empirical integral operator $T_{\mathbf{x}} : C(\mathscr{X}) \to C(\mathscr{X})$ is defined by

$$T_{\mathbf{x}} f(x) = \int_{\mathscr{X}} \frac{k(x, y)}{\sqrt{p_{\mathbf{x}}(x)}\sqrt{p_{\mathbf{x}}(y)}} f(y) dP_n(y)$$

$$= \frac{1}{n} \sum_{j=1}^n \frac{k(x, x_j)}{\sqrt{p_{\mathbf{x}}(x)}\sqrt{p_{\mathbf{x}}(x_j)}} f(x_j). \qquad (3)$$

* Corresponding author e-mail: gaowei@ynnu.edu.cn

Note that $p_\mathbf{x}(x_i) = \frac{1}{n}\sum_{j=1}^n k(x_i, x_j) = \frac{1}{n}(D_\mathbf{x})_{i,i}$ and $\frac{1}{\sqrt{p_\mathbf{x}(x_i)}} = \sqrt{n}(D_\mathbf{x}^{-\frac{1}{2}})_{i,i}$. Hence

$$[T_\mathbf{x} f(x_i)]_{i=1}^n = \left[\frac{1}{n}\sum_{j=1}^n \frac{k(x_i, x_j)}{\sqrt{p_\mathbf{x}(x_i)}\sqrt{p_\mathbf{x}(x_j)}} f(x_j)\right]_{i=1}^n$$

$$= D_\mathbf{x}^{-\frac{1}{2}} K_\mathbf{x} D_\mathbf{x}^{-\frac{1}{2}} [f(x_j)]_{j=1}^n.$$

It follows that if $(\lambda_n, u_n)$ is an eigenpair of $T_\mathbf{x}$, $[u_n(x_j)]_{j=1}^n$ is an eigenvector of the graph Laplacian $L_\mathbf{x}$ with eigenvalue $1 - \lambda_n$.

Convergence of the spectral clustering considered in this paper means that of $\{u_n\}_n$ (after normalization by $\|u_n\|_\infty = 1$) to an eigenfunction $u$ of T associated with a simple eigenvalue $\lambda \neq 0$ if $\{\lambda_n\}$ tends to $\lambda$. This convergence was verified when $k$ is a Gaussian similarity function $k(x, y) = \exp\{-\|x - y\|^2/\sigma^2\}$ on $\mathscr{X} \subset \mathbb{R}^n$ in [5] and when $k$ is a general Mercer kernel similarity function (with $p > 0$ but without positivity of $k$) in [3, ?], both with convergence rate $\mathscr{O}(\frac{1}{\sqrt{n}})$. A general estimate was also presented in [5], but the convergence associated with a general similarity function $k$ was not proved. The purpose of this paper is to verify the convergence of the spectral clustering associated with a general similarity function, and give quantitative estimates for the convergence. The convergence rates are stated in terms of the regularity of the similarity function $k$ and the capacity of the input space $\mathscr{X}$.

## 2 Ontology Measure and Ontology Mapping

The term of ontology comes from philosophy, it is used to describe the nature of things. Ontology similarity computation is widely used in medical science (see, for instance, [14] and [11]), biology science (see, for instance, [8]) and social science (see, for instance, [6]). Especially, in computer science, ontology is defined as a model for the sharing formal concepts have been applied in intelligent information integration, cooperative information systems, information retrieval, electronic commerce and knowledge management. As an effective model with hierarchical structure and semantics for the concepts, ontology technology has matured after nearly a decade of development. Now it has more systematic, comprehensive engineering theory, representation and construction tools. Particularly, in information retrieval, it has been used to compute semantic similarity (see [12]) and search extensions for concepts. Every vertex on an ontology graph represents a concept; A user searches for a concept $A$, will return similarities concepts of $A$ as search extensions to the user. Therefore, the quality of similarity functions play an important role in such applications. Moreover, ontology is also used in image retrieval (see, for instance, [7], [9], [10] and [13]).

When ontology is used in information retrieval, every vertex in ontology graph acts as a concept of ontology, the attraction of ontology graph can be used to measure the similarity of vertices. Let $G$ be a graph corresponding to ontology $O$, the goal of ontology similarity measure is to approach a similarity function which maps each pair of vertices to a real number. Choose the parameter $M \in \mathbb{R}_+$, the concepts $A$ and $B$ have high similarity if $Sim(A, B) > M$. The problem of ontology mapping can be described as follows: Let $G_1, G_2, \cdots, G_k$ be graphs corresponding to ontologies $O_1, O_2, \cdots, O_k$, respectively. Let $G = G_1 + G_2 + \cdots + G_k$. For every vertex $v \in V(G_i)$, where $1 \leq i \leq k$, the goal of ontology mapping is to find similarity vertices from $G - G_i$. Therefore, the ontology mapping problem is also an ontology similarity measure problem. Choose the parameter $M \in \mathbb{R}_+$, let $A, B$ be two concepts on ontology and $Sim(A, B) > M$, then return $B$ as retrieval expand when search concept $A$.

## 3 New Ontology Algorithm

Very recently, Gao et al. [15, 16] posed new algorithm for ontology measure and ontology mapping based on graph Laplacian. By using regularization framework of graph, the optimized function is obtained. Thus, all vertices in ontology graph are mapped into real numbers. The ontology measure and ontology mapping is obtained by comparing the difference of the corresponding values.

However, in many cases, ontology are changeable from time to time. There often new vertices added in ontology graph which regard as new concepts, and some old vertices may be deleted. In this situation, new vertices may be unlabelled. Thus, we need new tricks to deal with changeable ontology setting. One method is using semi-supervised learning (SSL). It is a graph Laplacian regularization for semi-supervised Laplacian eigenmap learning (see [17]).

The distribution on input label pairs $(v, y)$ is unknown, where $v \in \mathbb{R}^N$ and $y \in \mathbb{R}$. The aim of SSL is to learn a best predictor $f(v)$ for $y$ with few labeled and unlabeled examples. It should be considered SSL in a transductive setting, where we are given a labeled sample $(v_1, y_1), \cdots, (v_n, y_n)$ of $n$ labeled points, and an unlabeled sample $v_{n+1}, \cdots v_u$ of $u - n$ unlabeled points. The $u$ pairs $(v_i, y_i)$ are drawn i.i.d. from the source distribution. We use $V_L$ to denote the sequence of labeled vertices in ontology graphs, $V_U$ to denote the sequence of unlabeled vertices in ontology graphs, and $V = V_L \cup V_U$. The label sequences are denoted as $Y_L \in \mathbb{R}^n$, $Y_U \in \mathbb{R}^u$ and $Y \in \mathbb{R}^u$, respectively. Transductive SSL is to estimate $Y_U$ according to $V_L$, $Y_L$ and $V_U$. We write $f(\cdot)$ as a $u$-dimensional vector $f(V) = (f(v_1), \cdots, f(v_u)) \in \mathbb{R}^u$.

For a weighted ontology graph $G$ with $u$ vertices corresponding to the unlabeled and labeled vertices $v_1, \cdots, v_u$, and with edge weights $w_{ij}$. Let $W \in R^{u \times u}$ denote the weight matrix and its diagonal degree matrix $D$ with $D_{ii} = \sum_j w_{ij}$.

Let $\widehat{W} = D^{-1/2}WD^{-1/2}$ be a normalization of the weight matrix, and denote by $\widehat{D}$ its diagonal degree matrix $\widehat{D}_{ii} = \sum_j \widehat{w}_{ij}$. Then

$$\widehat{L}_r = I - \widehat{D}^{-1}\widehat{W}.$$

For a given weight matrix $\widehat{W}$, the unnormalized graph Laplacian defined as

$$\widehat{L}_u = \widehat{D} - \widehat{W}.$$

and the symmetric normalized graph Laplacian

$$\widehat{L}_s = \widehat{D}^{-1/2}\widehat{L}_u\widehat{D}^{-1/2} = I - \widehat{D}^{-1/2}\widehat{W}\widehat{D}^{-1/2}.$$

The eigenvectors of $\widehat{L}_r$ just as following the generalized eigenfunctions problem

$$\widehat{L}_u\pi_i = \lambda_i\widehat{D}\pi_i,$$

where $\widehat{L}_r = \widehat{D}^{-1}\widehat{L}_u$. The eigenvectors of $\widehat{L}_s$ and the right eigenvectors of $\widehat{L}_r$ have a one to one mapping. If $v_{r,i}$ is a right eigenvector of $\widehat{L}_r$ with eigenvalue $\lambda_i$, then $v_{s,i} = \widehat{D}^{1/2}v_{r,i}$ is an eigenvector of $\widehat{L}_s$ with the same eigenvalue $\lambda_i$.

Given the labeled and unlabeled data for certain ontology graph, and a parameter $t$, find the leading $t$ right eigenvectors $v_{r,1}, \cdots, v_{r,t}$ of $\widehat{L}_r$, with the smallest eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_t$. By the mapping $v_i \rightarrow (v_{r,1}(v_i), v_{r,2}(v_i), \ldots, v_{r,k}(v_i))$, where $v_{r,j}(v_i)$ is the coordinate of $v_{r,j}$, it can be performed an ordinary (unregularized) least squares regression in the $t$-dimensional space. Especially, the least squares predictor

$$\widehat{\beta} = \text{argmin}_\beta \sum_{i=1}^n (y_i - \sum_{j=1}^k \beta_j v_{r,j}(v_i))^2$$

and predict

$$\widehat{f}(v_i) = \sum_{j=1}^k \widehat{\beta}_j v_{r,j}(v_i).$$

The semi-supervised eigermap learning algorithm can be used in ontology concepts similarity measurement. The basic idea is: via the semi-supervised eigermap learning algorithmthe ontology graph is mapped into a line consists of real numbers. The similarity between two concepts then can be measured by comparing the difference between their corresponding real numbers.

Algorithm for ontology similarity measure: For $v \in V(G)$, we use one of the following methods to obtain the similar vertices and return the outcome to the users. Choose a parameter $M$return set $\{u \in V(G), |f(u) - f(v)| \leq M\}$.

Algorithm for ontology mapping: For $v \in V(G_i)$, where $1 \leq i \leq k$, we use one of following methods to obtain the similar vertices and return the outcome to the users. Choose a parameter $M$, return set $\{u \in V(G - G_i), |f(u) - f(v)| \leq M\}$.

## 4 Experiments

To connect ontology to this semi-supervised eigermap learning algorithm, we should use a vector to express each vertex of information. This vector contains the information of name, instance, attribute and structure of the vertex, where the instance of vertex is the set of its reachable vertices in the directed ontology graph.

The first experiment concerns ontology similarity measurement is described as follows. In this experiment we use computer ontology $O_1$ which was constructed in [20]. The goal of the algorithm is to map the vertices on the graph into a line consists of real numbers. The similarity between two concepts then can be measured by comparing the difference between their corresponding real numbers. Thusthe similarities we obtained are indirect similarity measure not the direct ones. We use $P@N$ ( Precision Ratiosee [21]) to measure the equality of the experiment. Firstthe expert gives the first $N$ concepts for every vertex on the ontology graphthen we obtain the first $N$ concepts for every vertex on ontology graph by the algorithm and compute the precision ratio.

The experiment shows that, $P@1$ Precision Ratio is 0.39, $P@3$ Precision Ratio is 0.55, $P@5$ Precision Ratio is 0.70. Thus the proposed algorithm has high efficiency.

For the second experimentwe use another Computer Ontologies $O_2$ which constructed in [20] shows $O_2$. The goal of the algorithm is to map the vertices on $G = G_1 + G_2$ into a line consists of real numbers. We also use P@N Precision Ratio to measure the equality of experiment.

The experiment shows that, $P@1$ Precision Ratio is 0.33, $P@3$ Precision Ratio is 0.48, $P@5$ Precision Ratio is 0.63. Thus the algorithm has high efficiency.

## 5 Conclusion

In this paper, we give a new algorithm for measuring the ontology similarity and ontology mapping using semi-supervised eigermap learning. The new algorithms have high quality according to the experiments above.

## Acknowledgement

## References

[1] M. Belkin and P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. **15**, 1373–1396 (2003).

Information Sciences Letters, Vol. 2 [2013], Iss. 2, Art. 6

92　　Bo Peng, Tian-wei Xu, Zhen Li, Wei Gao: Ontology Algorithm Based on...

[2] F. Chung, Spectral Graph Theory, CBMS Regional Conference Series in Mathematics **92**, AMS, (1996).

[3] V. Koltchinskii and E. Giné, Random matrix approximation of spectra of integral operators, Bernoulli **6**, 113–167 (2000).

[4] S. Smale and D. X. Zhou, Geometry on probability spaces, Constr. Approx. **30**, 311-323 (2009).

[5] U. von Luxburg, M. Belkin, and O. Bousquet, Consistency of spectral clustering, Ann. Stat. **36**, 555–586 (2008).

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

[7] B. Hu, S. Dasmahapatra, P. Lewis, and N. Shadbolt, Ontology-based medical image annotation with description logics, 15th IEEE International Conference on Tools with Artificial Intelligence, 77-82 (2003).

[8] P. Lambrix, A. Edberg, Evaluation of ontology tools in bioinformatics. In: Paci.c Symposium on Biocomputing, 529-600 (2003).

[9] S. Liu , L. Chia, and S. Chan, Ontology for naturescene image retrieval, In on the move to meaningful Internet systems, CoopIS, DOA, and ODBASE, 1050-1061 (2004).

[10] J. Opitz, B. Parsia, and U. Sattler, Using Ontologies for Medical Image Retrieval- An Experiment, Proceedings 5th Int. Workshop on OWL: Experiences and Directions, **529**, 23-24 (2009).

[11] I. Spasic, S. Ananiadou, J. McNaught, and A. Kumar, Text mining and ontologies in biomedicine: making sense of raw text, Briefings in bioinformatics, **6**, 239-251 (2006).

[12] X. Su and J. Gulla, Semantic enrichment for ontology mapping, The 9th International Conference to Information Systems (NLDB), 217-228 (2004).

[13] H. Wang, S. Liu, and L. Chia, Does ontology help in image retrieval?-A comparison between keyword, text ontology and multi-modality ontology approaches. MM06, October, Santa Barbara, California, USA. ACM, 109-112 (2006).

[14] P. Mork, A. Philip, and Bernstein, Adapting a generic match algorithm to align ontologies of human anatomy. In: 20th International Conf. on Data Engineering. IEEE, Los Alamitos. **20** 787-790 (2004).

[15] W. Gao, L. Liang, and Y. Zhang, Ontology Concept Similarity Computation Based on Graph Learning, Journal of Southwest China Normal University (Natural Science Edition) **36** 64-67 (2011).

[16] W. Gao, L. Zhu, and L. Liang,Ontology mapping algorithm based on the regularization framework of graph, Journal of southwest university (Natural science edition) **34**, 118-121 (2012).

[17] X. Zhu, Z. Ghahramani, and J. Lafferty, Semi-Supervised Learning Using Gaussian Fields and Harmonic Function. In The Twentieth International Conference on Machine Learning, 2003.

[18] D. Zhou, Olivier Bousquet, Thomas Navin Lal, Ja- son Weston, and Bernhard Scholkopf, Learning with Local and Global Consistency. In Sebastian Thrun, Lawrence Saul, and Bernhard Scholkopf, editors, Advances in Neural Information Processing Systems, Cambridge, MA, 2004. MIT Press.

[19] M. Belkin, I. Matveeva, and P. Niyogi, Regularization and Semi-supervised Learning on Large Graphs. In John Shawe-Taylor and Yoram Singer, editors, COLT, Lecture Notes in Computer Science,Springer, **3120**, 624-638 (2004).

[20] X. Huang, T. Xu, W. Gao, Z. Jia, Ontology similarity measure and ontology mapping via fast ranking method, International Journal of Applied Physics and Mathematics, 2011, **1**, 54-59 (2011).

[21] N. Craswell, D. Hawking, Overview of the TREC 2003 web track, in Proc. The 12th Text Retrieval Conference. Gaithersburg.