# Theoretical Characteristics of Ontology Learning Algorithm in Multi-dividing Setting

Linli Zhu, Weige Tao, Xiaozhong Min and Wei Gao

*Abstract*—**Ontology, as a useful tool, has been widely applied in various fields, and ontology concept similarity calculation is an essential problem in these application algorithms. A recent method to get similarity between vertices on ontology is not by pairwise computation but based on a function which maps ontology graph into a line and maps every vertex in graph into a real-value, the similarity is measured by the difference of their corresponding scores. The multi-dividing method is suitable for ontology problem and plays a key role in achieving this. Such ontology function is given by learning a training sample which contains a subset of vertices with $k$ classes from ontology graph. In this paper, we propose a new multi-dividing ontology algorithm framework, which is designed to avoid the choice of loss function. Meanwhile, there is such a vertex selection policy in new multi-dividing ontology algorithm that it guarantees that the new algorithm can be employed for an ontology graph with its structure rather than a tree. We provide some theoretical characteristics of the new multi-dividing ontology algorithm, and show that the new algorithm is convergent.**

*Index Terms*—**Ontology, multi-dividing, ROC optimization, AUC criterion, VC major class.**

## I. Introduction

THE term of "Ontology", deriving from philosophy, is used to describe the nature of things. In computer science, ontology is defined as a shared conceptual model, which has been applied in intelligent information integration, collaboration, information systems, information retrieval, e-commerce, knowledge management and image retrieval. As an effective conceptual semantic model, ontology technology has been widely employed in many other areas such as social science (for instance, see Qiu and Lou [1]), biology medicine (for instance, see Arsene et al., [2]), and geography science (for instance, see Arsene et al., [2]).

In information retrieval, ontology has been used to compute semantic similarity (for instance, see Su and Gulla [4]) and search extensions for concepts. Every vertex on an ontology graph represents a concept; a user, searching for a concept $A$, will return similarities concepts of $A$ as search extensions. 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 $Sim(A,B) > M$. Let $A, B$ be two concepts on ontology and if $Sim(A,B) > M$, then return $B$ as retrieval expand when search concept $A$. Therefore, the quality of similarity functions plays an important role in such applications.

Manuscript received Janurary 6, 2016; revised February 20, 2016. This work was supported in part by the National Natural Science Foundation of China(NO.61142007), Jiangsu Province Science and Technology PlanIndustry-University-Research Innovation Funds – Prospective Research ProjectNO.BY2014038-09and the general program of Jiangsu University of Technology in China (NO.KYY14013).

Linli Zhu is with the Department of Computer Engineering, Jiangsu University of Technology, Changzhou, China.(email:zhulinli@jsut.edu.cn).

Weige Tao is with the Department of Electricity and Information Engineering, Jiangsu University of Technology, Changzhou, China.

Xiaozhong Min is with the Department of Computer Engineering, Jiangsu University of Technology, Changzhou, China.

Wei Gao is with the Department of Information and Technology, Yunnan Normal University, Kunming,China.

Recent years have witnessed some effective technologies for ontology similarity measurement. Zhu et al., [5] proposed a new algorithm to get a similarity matrix in terms of learning the optimal similarity kernel function for ontology applications. Zhu and Gao [6] presented a new optimization model for ontology similarity measurement and ontology mapping in multi-dividing setting such that the similarity function was designed from the idea of partial AUC criterion. Gao et al., [7] obtained the fast ontology algorithm for standard ontology SVM by virtue of infinite push multi-dividing ontology algorithm. Gao et al., [8] deduced an ontology optimize algorithm for ontology sparse vector learning using gradient descent, and a fast rate version of algorithm was also given. Gao and Gao [9] inferred a sparse vector learning algorithm for ontology similarity measure and ontology mapping in view of accelerated first-order technology. Lan [10] explored the learning theory approach for ontology similarity computation in a setting when the ontology graph is a tree. It forms $k$ branches in the tree and any two vertices belong to different branches have no edge between them. The concepts in the same branch of the tree should have higher similarity, compared with concepts in different branches. Using the multi-dividing algorithm in which the $k$ parts correspond to the $k$ classes of vertices of $k$ rates. The rate values of all classes are decided by experts. Then, a vertex in a rate $a$ has larger value than any vertex in rate $b$ (where $1 \leq a < b \leq k$) under ontology function $f$. At last, the similarity between two vertices is judged by the difference between two real numbers which they correspond to. Thus, the multi-dividing algorithm is reasonable to learn a score function for an ontology graph with a tree structure. More ontology algorithms can be referred to Gao and Shi [11] and Lan et al., [12].

Furthermore, in [13], Gao and Xu studied the uniform stability of multi-dividing ontology algorithm and obtained the generalization bounds for stable multi-dividing ontology algorithms. Gao et al., [14] researched the strong and weak stability of multi-dividing ontology algorithm. Gao and Xu [15] learned some characters for such ontology algorithm. Other analysis for ontology algorithm can be refered to [16], [17], [18].

In this paper, we still focus on the theoretical analysis of multi-dividing ontology algorithm, but the new multi-dividing framework is reconsidered by us from a new perspective. We obtain ontology score function under the new multi-dividing framework which is regardless of the loss function, and propose a vertex partitioning strategy to solve the problem of ontology graph structure restrictions for multi-dividing setting. We focus on the theoretical analysis of the new framework. The convergence of new multi-dividing ontology algorithm is proved.

The organization of this paper is as follows: we describe the multi-dividing ontology problem in Section 2, notions for new multi-dividing ontology algorithm setting are defined in Section 3. Using these notions and new idea, we derive some theoretical results for new multi-dividing ontology algorithm in Section 4.

## II. The multi-dividing method for ontology

Let $V \subset \mathbb{R}^q$ ($q \geq 1$) denote an input space (or the instance space) for ontology graph, and the vertices (or, instances) in $V$ are drawn randomly and independently according to some (unknown) distribution $\mathcal{D}$. Given a training set $S = \{v_1, \cdots, v_n\}$ of size $n$ in $V$, the goal of ontology learning algorithms is to obtain a score function $f : V \to \mathbb{R}$, which assigns a score to each vertex, and ranks all the instances according to their scores. The multi-dividing ontology problem is a special kind of ontology learning problem in which vertices come from $k$ categories and the learner is given examples of vertices labeled as there $k$ classes.

Formally, the settings of multi-dividing ontology problems can be described as follows. There is an instance space $V$ from which vertices are drawn, and the learner is given a training sample $(S_1, S_2, \cdots, S_k) \in V^{n_1} \times V^{n_2} \times \cdots \times V^{n_k}$ consisting of a sequence of training sample $S_a = (v_1^a, \cdots, v_{n_a}^a)$ ($1 \leq a \leq k$). The goal is to learn from these samples a real-valued ontology score function $f : V \to \mathbb{R}$ that orders the future $S_a$ vertices rank higher than $S_{a'}$ where $a < a'$, that is, $f$ is considered to score a vertex $v$ higher than a vertex $v'$ if $f(v) > f(v')$. Let $\prec_f$ be a preorder on $V$ such that for any $(v, v') \in V^2$, $v \prec_f v'$ if and only if $f(v) \leq f(v')$. We assume that instances in each $S_a$ are drawn randomly and independently according to some (unknown) distribution $\mathcal{D}_a$ on the instance space $V$ respectively. Denote $Y$ be a label of $V$, which indicates its classification information. We use pair $(\mu, \eta)$ to describe the probability measure on the underlying space, where $\mu$ implies the marginal distribution of $V$ and $\eta^{a,b}(v) = \mathbb{P}\{Y = a | V = v, Y \in \{a, b\}\}$, $v \in V$ and $a \in \{1, \cdots, k\}$ denotes the posterior probability. Without loss of generality, we always assume that $V$ coincides with the support of $\mu$. If not particularly specified, $\eta$ refers to established for all $(a, b)$.

The ontology loss function $l : \mathbb{R}^V \times V \times V \to \mathbb{R}^+ \cup \{0\}$ is used to punish the inconsistent situation which sgn$(f(v) - f(v'))$ is not coincide with their category relationships, where sgn$(u)$=$\begin{cases} 1, & u > 0 \\ 0, & u = 0 \\ -1, & u < 0 \end{cases}$. $l$ also satisfies that $l(f, v, v')$ is a non-negative real number and symmetric with respect to $v$ and $v'$. As an example, one common loss function called $\gamma$ ontology loss for $\gamma > 0$ defined as

$$l_\gamma(f, v, v') = \begin{cases} 1, & \text{if } f(v) - f(v') \leq 0 \\ 1 - \frac{f(v) - f(v')}{\gamma}, & \text{if } 0 < f(v) - f(v') < \gamma \\ 0, & \text{if } f(v) - f(v') \geq \gamma \end{cases}$$

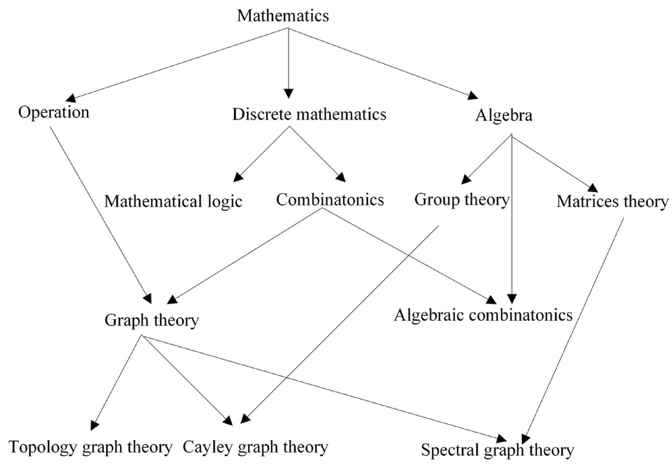


Fig. 1. Mathematics Ontology

The quality of the ontology score function is measured by the expected $l$-error:

$$R_l(f) = \tag{1}$$

$$\frac{1}{\sum_{a=1}^{k-1} \sum_{b=a+1}^{k}} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} n_a n_b \mathbb{E}_{v \sim \mathcal{D}_a, v' \sim \mathcal{D}_b} \{l(f, v, v')\}.$$

However, computation model (1) cannot be calculated directly since the distribution $\mathcal{D}_a$ for $a = 1, \cdots, k$ are unknown. Instead, we use empirical l-error to measure ontology function:

$$\begin{aligned} & \widehat{R}_l(f; S_1, \cdots, S_k) \qquad (2) \\ = & \frac{1}{\sum_{a=1}^{k-1} \sum_{b=a+1}^{k}} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \sum_{i:v_i^a} \sum_{j:v_j^b} l(f, v, v'). \end{aligned}$$

Calculation model (2), however, has an obvious drawback which is that it depends on the selection of the ontology loss function $l$. Different loss functions lead to different optimal ontology score functions, and often the distinction of performances for different optimal ontology score functions are extensively large. At the same time, the theoretical analysis in [13], [14], [15] showed that the theoretical characters of different ontology score functions given by different ontology loss functions have big gap due to their different consistent continuity, differentiability, smoothness, and other mathematical properties. It inspires us to consider about obtaining an optimal ontology function from the nature of the multi-dividing algorithm, which avoids the choice of ontology loss function.

From [13], [14], [15], the use of the multi-dividing ontology algorithm is based on the assumption that the ontology graph is a tree, i.e., an acyclic graph. Under this premise, all vertices on given ontology graph except for the top vertex can be uniquely classified. However, in certain application areas, there exists a cycle in ontology graph. For example, see the “Mathematics ontology” in Fig. 1.

In this ontology graph, vertices “Cayley graph theory” and “Spectral graph theory” belong to cross-discipline, and they are associated with more than one branch of mathematics. If we apply the multi-dividing method in this ontology, then $k = 3$, and all vertices can be divided into three parts: “Operation”, “Discrete mathematics”, “Algebra”. However, these interdisciplinary vertices are difficult to determine which part they belong to.

## III. Definitions

In order to solve the first problem, our proposal in this paper is to evaluate the quality of ontology function by means of its ROC (Receiver Operating Characteristic) curve. The trick of ROC curve has been widely used in classification, regression and ranking (for instance, see [19], [20], [21]). Here, we first define the multi-dividing version of ROC curve for our ontology algorithm use. In the following text, all the ontology score functions considered are connected with ontology multi-dividing algorithm setting, and we hereby make no specific statement again.

*Definition 1:* Let $f$ be an ontology score function. The ROC curve of $f$ is given by

$$\begin{aligned} t \in \mathbb{R} \quad \rightarrow \quad & (\mathbb{P}\{f(V) > t | Y = 1\}, \cdots, \\ & \mathbb{P}\{f(V) > t | Y = k\}) \in [0,1]^k. \end{aligned}$$

When the jump occurs in ROC curve, line segments are employed to connect the corresponding extremities of curve. From this point of view, ROC curve of ontology score function $f$ can be regarded as the graph of a continuous mapping $\alpha \in [0,1] \rightarrow ROC(f, \alpha)$.

In a certain sense, the ROC curve implies a partial order for ontology score function set: for all pairs of ontology score function $f_1$ and $f_2$, if ROC$(f_1, \alpha) \leq$ ROC$(f_2, \alpha)$ holds for any $\alpha \in [0,1]$, then we say that $f_2$ is more accurate than $f_1$. In view of standard Neyman-Pearson argument, the most accurate ontology score function can be established from increasing transform of the function, which is similar to the conditional probability function depending on an affine transformation.

*Definition 2:* (Optimal ontology score function) Let $F^*$ be the set of ontology score functions. If for any $f^* \in F^*$, and for each pair $(v, v') \in V^2$, $\eta(v) < \eta(v') \Rightarrow f^*(v) < f^*(v')$, then we call $F^*$ is the optimal ontology score function set.

In fact, the elements of $f^*$ are optimizers of the ROC curve. Moreover, if we assume the random variable $\eta(V)$ to be continuous, then $F^*$ coincides with the set of strictly increasing transforms of $\eta$. The performance of a candidate ontology score function $f$ is usually measured by AUC (Area Under the ROC Curve), which can be regarded as a summary of the ROC curve. We use the following version of definition for AUC, which corresponds to multi-dividing ontology algorithm.

*Definition 3:* Let $f$ be an ontology score function. The AUC is the function defined as:

$$\begin{aligned} & \mathrm{AUC}(f) \\ = & \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\mathbb{P}\{f(V_a) > f(V_b) | (Y_a, Y_b) = (a,b)\} \\ & + \frac{1}{2} \mathbb{P}\{f(V_a) = f(V_b) | (Y_a, Y_b) = (a,b)\}\}. \end{aligned}$$

where $(V_a, Y_a)$ and $(V_b, Y_b)$ denote two independent copies of the pair $(V, Y)$, for any ontology score function $f$.

This function gives a total order on the set of ontology score functions and AUC$(f)$ coincides with $\int_0^1$ROC$(f, \alpha)$. We denote the optimal curve and the corresponding (maximum) value for the AUC criterion by ROC$^*$ =ROC$(f^*, \cdot)$ and AUC$^*$=AUC$(f^*)$ respectively, where $f^* \in F^*$. The statistical counterparts of ROC$(f, \cdot)$ and AUC$(f)$ denoted by $\widehat{\mathrm{ROC}}(f, \cdot)$ and $\widehat{\mathrm{AUC}}(f)$, rely heavily on sampling data which are obtained by replacing the class distributions by their empirical versions in the definitions. We focus on a particular subclass of ontology score function in this paper.

*Definition 4:* ( Piecewise constant ontology score function) An ontology score function $f$ is piecewise constant if there exists a finite partition $P$ of $V$ such that for all $C \in P$, there exists a constant $k_C \in \mathbb{R}$ such that for any vertex $v \in C$, $f(v) = k_C$.

However, Definition 4 does not provide a unique characterization of the underlying partition. The partition $P$ is called minimal if, for any two of its elements $C \neq C'$, we have $k_C \neq k_{C'}$. The ontology score function conveys an ordering on the cells of the minimal partition.

*Definition 5:* Let $f$ be an ontology score function and $P$ be its associated minimal partition. The ontology score function induces an order $\prec_f$ over the cells of the partition. For a given cell $C \in P$, its order $R_{\prec_f} \in \{1, \cdots, |P|\}$ which is affected by the $\prec_f$ over the elements of the partition. By convention, we set rank 1 to correspond to the highest score.

The ontology score functions considered in this paper result from a collection of piecewise constant ontology score functions. Since each of these ontology score function is related to a possibly different partition, we must consider a collection of partitions of $V$. In our proposal, good ontology score function is defined on the least fine subpartition of this collection of partitions.

*Definition 6:* Consider a collection of $k$ partitions of $V$ denoted by $P_b$, $b = 1, \cdot, k$. A subpartition of this collection is a partition $P_b$ made of nonempty subsets $C \subset V$ such that for each $C \in P_k$, there exists $(C_1, \cdots, C_k) \in P_1 \times \cdots \times P_k$ satisfying

$$C \subseteq \bigcup_{b=1}^{k} C_b.$$

We denote $P_k^* = \cap_{b \leq k} P_b$.

An obvious fact is that $P_k$ is a subpartition of any of the $P_b$'s. From the point that any partition $P$ which is a subpartition of $P_b$ for any $b \in \{1, \cdots, k\}$, the largest one is just a subpartition of $P_k$. This gives us the theoretical basis of how to divide the vertex when ontology graph structure is not a tree. Considering a collection of piecewise constant ontology score function $f_b$, $b = 1, \cdots, k$, and $P_b$ is denoted as their corresponding minimal partitions. Naturally, every ontology score function $f_b$ induces a ordering $\prec_b^*$ on the partition $P_k^*$. For any $(C, C') \in P_k^{*2}$, $(C_b, C_b^{'}) \in P_b^2$ and $C \times C' \subset C_b \times C_b^{'}$, we infer that $C \prec_b^* C'$ if and only if $C_b \prec_b^* C_b^{'}$.

The collection of ontology score functions results in a collection of $k$ orderings on $P_k^*$. In voting theory, such a collection is named a profile. Next, we will consider the case of piecewise constant ontology score functions which is based on the definition of the probabilistic Kendall $\tau$. We use more curly notation for the preorder relation $\prec_f$ on $V$ with the following restriction: for any $C, C' \in P$, $v \in C$ and $v' \in C'$, we get $v \prec_f v'$ if and only if $C \prec_f C'$. Equivalent to that, $f(v) \leq f(v')$ for any $v \in C$ and $v' \in C'$. We should introduce a measure of similarity for preorders on V induced by ontology score functions $f_1$ and $f_2$.

Recall that the probabilistic Kendall $\tau$ for two random variables $(Z_1, Z_2)$ on the same probability space is defined as

$$\tau(Z_1, Z_2) = 1 - 2d_\tau(Z_1, Z_2),$$

where

$$\begin{aligned} d_\tau(Z_1, Z_2) &= \mathbb{P}\{(Z_1 - Z_1^{'}) \cdot (Z_2 - Z_2^{'}) < 0\} \\ &+ \frac{1}{2}\mathbb{P}\{Z_1 = Z_1^{'}, Z_2 \neq Z_2^{'}\} \\ &+ \frac{1}{2}\mathbb{P}\{Z_1 \neq Z_1^{'}, Z_2 = Z_2^{'}\} \end{aligned}$$

and $(Z_1^{'}, Z_2^{'})$ is an independent copy of the pair $(Z_1, Z_2)$.

## IV. Theoretical analysis

In terms of the definition of probabilistic Kendall $\tau$, we derive that the Kendall $\tau$ for the pair $(f(V), Y)$ is related to AUC($f$). This fact implies that: if we use natation $p_a^{a,b} = \mathbb{P}\{Y = a | Y \in \{a, b\}\}$, then for any real-valued ontology score function $f$, we infer:

$$\begin{aligned} &\frac{1}{2}(1 - \tau(f(V), Y)) \\ = &\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} 2p_a^{a,b}p_b^{a,b}(1 - \mathrm{AUC}(f)) \\ &+ \frac{1}{2}\mathbb{P}\{f(V) \neq f(V'), Y = Y'\}. \end{aligned}$$

For two given ontology score functions $f_1$ and $f_2$ and considering the probabilistic Kendall tau for random variables $f_1(V)$ and $f_2(V)$, we denote $d_V(f_1, f_2) = d_\tau(f_1(V), f_2(V))$. It simply manifests that $d_V$ implies a distance between $\prec_{f_1}$ and $\prec_{f_2}$ induced by ontology score functions $f_1$ and $f_2$ on the vertex set $V$. The following lemma shows that the deviation between ontology score functions in terms of AUC is dominated by a quantity involving the probabilistic agreement based on Kendall $\tau$.

*Lemma 1:* Assume $p_a \in (0, 1)$ for $a = 1, \cdots, k$. For any ontology score function $f_1$ and $f_2$ on $V$, we obtain

$$\begin{aligned} &|\mathrm{AUC}(f_1) - \mathrm{AUC}(f_2)| \\ \leq &\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \frac{d_V(f_1, f_2)}{2p_a^{a,b}p_b^{a,b}} \\ = &\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \frac{1 - \tau_V(f_1, f_2)}{4p_a^{a,b}p_b^{a,b}}. \end{aligned}$$

**Proof.** In view of $\tau_V(f_1, f_2) = 1 - 2d_V(f_1, f_2)$, where $d_V(f_1, f_2)$ is given by:

$$\begin{aligned} &\mathbb{P}\{(f_1(V) - f_1(V')) \cdot (f_2(V) - f_2(V')) < 0\} \\ + &\frac{1}{2}\mathbb{P}\{f_1(V) = f_1(v'), f_2(V) \neq f_2(v')\} \\ + &\frac{1}{2}\mathbb{P}\{f_1(V) \neq f_1(v'), f_2(V) = f_2(v')\}. \end{aligned}$$

For any $f$, AUC($s$) can be rewritten as:

$$\begin{aligned} &\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \frac{\mathbb{P}\{(f(V) - f(V'))(Y - Y') > 0\}}{2p_a^{a,b}p_b^{a,b}} \\ + &\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \frac{\mathbb{P}\{f(V) = f(V'), Y = Y'\}}{4p_a^{a,b}p_b^{a,b}}. \end{aligned}$$

Using Jensen's inequality, we easily obtain that $\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} 2p_a^{a,b}p_b^{a,b}|\mathrm{AUC}(f_1) - \mathrm{AUC}(f_2)|$ is bounded by the expectation of the random variable

$$\begin{aligned} &\Pi\mathbb{P}\{(f_1(V) - f_1(V')) \cdot (f_2(V) - f_2(V')) < 0\} \\ + &\frac{1}{2}\Pi\mathbb{P}\{f_1(V) = f_1(v'), f_2(V) \neq f_2(v')\} \\ + &\frac{1}{2}\Pi\mathbb{P}\{f_1(V) \neq f_1(v'), f_2(V) = f_2(v')\} \end{aligned}$$

which is equal to $d_V(f_1, f_2) = \frac{1 - \tau_V(f_1, f_2)}{2}$. □

What we need to pay special attention to is the fact that the converse inequality does not hold in general. Ontology score functions with same AUC may lead to different orders. To our delight, the following lemma ensures that under assumption of the noise condition, an ontology score function with a nearly optimal AUC is approximate to the optimal ontology score functions in the sense of Kendall $\tau$.

*Lemma 2:* Suppose that the random variable $\eta(V)$ is continuous and that there exists $c < \infty$ and $d \in (0, 1)$ satisfy

$$\mathbb{E}[|\eta(V) - \eta(v)|^{-d}] \leq c \tag{3}$$

for each $v \in V$. Then, for any ontology score function $f$ and any optimal ontology score function $f^* \in F^*$, we get

$$1 - \tau_V(f^*, f) \leq C(\mathrm{AUC}^* - \mathrm{AUC}(f))^{\frac{d}{1+d}},$$

where $C$ is constant with

$$C = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} 3c^{\frac{1}{1+d}}(2p_a^{a,b}p_b^{a,b})^{\frac{d}{1+d}}$$

**Proof.** In terms of the fact that, for any $f \in F$, the AUC deficit

$$\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} 2p_a^{a,b}p_b^{a,b}|\mathrm{AUC}(f_1) - \mathrm{AUC}(f_2)|$$

can be rewritten as

$$\begin{aligned} &\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbb{P}\{f(V) = f(V'), (Y, Y') = (a, b)\} \\ + &\mathbb{E}[|\eta^{a,b}(V) - \eta^{a,b}(V')| \cdot \Pi\{(V, V') \in \Gamma_f^{a,b}\}], \end{aligned}$$

where

$$\Gamma_f^{a,b} = \{(v, v') \in V^2 : (f(v) - f(v')) \cdot (\eta^{a,b}(v) - \eta^{a,b}(v')) < 0\}.$$

Combining with Holder inequality and noise condition (3), we infer that $\mathbb{P}\{(V, V') \in \Gamma_f^{a,b}\}$ is bounded by

$$(\mathbb{E}[|\eta^{a,b}(V) - \eta^{a,b}(V')| \cdot \Pi\{(V, V') \in \Gamma_f^{a,b}\}])^{\frac{d}{1+d}} \times c^{\frac{1}{1+d}}.$$

Thus, for every $f^* \in F^*$, we have

$$\begin{aligned} d_V(\prec_f, \prec_{f^*}) &= \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbb{P}\{(V, V') \in \Gamma_f^{a,b}\} \\ &+ \frac{1}{2}\mathbb{P}\{f(V) = f(V')\}. \end{aligned}$$

Since $\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} p_a^{a,b}p_b^{a,b}\mathbb{P}\{f(V) = f(V'), (Y, Y') = (a, b)\}$ can be represented as

$$\begin{aligned} & \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbb{E}[\Pi\{f(V) = f(V')\} \\ & \cdot\eta^{a,b}(V')(1-\eta^{a,b}(V))] \\ = & \frac{1}{2}\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbb{E}[\Pi\{f(V) = f(V')\} \\ & \cdot(\eta^{a,b}(V') + \eta^{a,b}(V) - 2(\eta^{a,b}(V')\eta^{a,b}(V)))], \end{aligned}$$

which can be easily verified to be larger than $\frac{1}{2}\sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbb{E}[\Pi\{f(V) = f(V')\} \cdot |\eta^{a,b}(V') - \eta^{a,b}(V)|]$. With the similar discussion demonstrated above, we get that $\mathbb{P}\{f(V) = f(V')\}$ is bounded by

$$\begin{aligned} & \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} (\mathbb{E}[|\eta^{a,b}(V) - \eta^{a,b}(V')| \\ \cdot \quad & \Pi\{f(V) = f(V')\}])^{\frac{d}{1+d}} \times c^{\frac{1}{1+d}}. \end{aligned}$$

From what we argued above, we deduce the desired result. □

The following lemma reveals the link between a specific notion of distance between $\prec_{f_1}$ and $\prec_{f_2}$ on $P$ and the Kendall $\tau$ distance between preorders on vertex set $V$ induced by piecewise constant ontology score functions $f_1$ and $f_2$. This result is obtained directly from the definition, thus we omitted the certification process.

*Lemma 3:* Let $f_1$ and $f_2$ be two piecewise constant ontology score functions. We obtain

$$d_V(f_1, f_2) = 2\sum_{1\le r<l\le R} \mu(C_r)\mu(C_l)U_{r,l}(\prec_{f_1}, \prec_{f_2}), \quad (4)$$

where

$$\begin{aligned} & U_{r,l}(\prec, \prec') \\ = & \Pi\{(R_{\prec}(C_r) - R_{\prec}(C_l)) \cdot (R_{\prec'}(C_r) - R_{\prec'}(C_l)) < 0\} \\ & +\frac{1}{2}\Pi\{R_{\prec}(C_r) = R_{\prec}(C_l), R_{\prec'}(C_r) \neq R_{\prec'}(C_l)\} \\ & +\frac{1}{2}\Pi\{R_{\prec}(C_r) \neq R_{\prec}(C_l), R_{\prec'}(C_r) = R_{\prec'}(C_l)\} \end{aligned}$$

for any two orderings $\prec$, $\prec'$ on a partition of cells $\{C_r : r = 1, \cdots, R\}$.

In the right side of (4), $U_{r,l}(\prec_{f_1}, \prec_{f_2})$ equals to 1 if the cells $C_r$ and $C_l$ are not classified in the same order by ontology score functions $f_1$ and $f_2$, to $\frac{1}{2}$ if they are tied for one ordering but not for the other, and to 0 for other cases. In conclusion, the agreement $\tau_V(f_1, f_2)$ can be regarded as a weighted version of the rate of accordant pairs for the cells of $P$, which is measured by Kendall $\tau$. By substituting the values of $\mu(C_r)$ for their empirical counterparts in (4), we can get the statistical version of $\tau_V(f_1, f_2)$. Thus, let

$$\widehat{\tau}_V(f_1, f_2) = 1 - 2\widehat{d}_V(f_1, f_2), \quad (5)$$

where $\widehat{d}_V(f_1, f_2) = 2/(n(n-1))\sum_{i<j} K(V_i, V_j)$ is a $U$-statistic with degree 2, and $K(\cdot, \cdot)$ is symmetric kernel defined as follows:

$$\begin{aligned} K(v, v') = & \Pi\{(f_1(v) - f_1(v')) \cdot (f_2(v) - f_2(v')) < 0\} \\ & +\frac{1}{2}\Pi\{f_1(v) = f_1(v')\}\Pi\{f_2(v) \neq f_2(v')\} \\ & +\frac{1}{2}\Pi\{f_1(v) \neq f_1(v')\}\Pi\{f_2(v) = f_2(v')\}. \end{aligned}$$

The trick we consider in this paper heavily depends on the median procedure, which regards the family of metric procedures (see Barthelemy and Montjardet [22] for more details). Let $d(\cdot, \cdot)$ be dissimilarity measure or certain metric on the set of orderings on a finite set $Z$. In terms of its definition, a median ordering with respect to $d$ among a profile $\Pi = \{r : 1 \le r \le R\}$ is just ordering $\prec_{\rm med}$ on $Z$ satisfies

$$\Delta_{\Pi}(\prec_{\rm med}) = \min_{\prec:\ \text{ordering on } Z} \Delta_{\Pi}(\prec),$$

which implies that minimizing the sum $\Delta_{\Pi}(\prec) = \sum_{r=1}^{R} d(\prec, \prec_r)$ over the set $R(Z)$ of all orderings $\prec$ on set $Z$.

Let $N < \infty$ be cardinality of $Z$. We derive that there exists

$$|R(Z)| = \sum_{r=1}^{N}(-1)^r\sum_{m=1}^{r}(-1)^m\binom{r}{m}m^N$$

possible orderings on $Z$ and in worst cases in computation of median orderings of this combinatorial optimization problem is NP-hard.

When we consider preorders with respect to vertex set $V$ of infinite cardinality, it becomes more difficult to define such notion. Fixed a pseudo-metric such as $d_\tau$ and ontology score functions $f_1, \cdots, f_k$ on $V$, we are not sure whether there exists $\overline{f}$ in $F$ with $\sum_{k=1}^{K} d_\tau(\overline{f}, f_k) = \min_f \sum_{k=1}^{K} d_\tau(f, f_k)$. However, when it comes to piecewise constant ontology score functions with corresponding finite subpartition $P$ on $V$, there is one-to-one correspondence between the corresponding preorders and orderings on $P$. In this way, the minimum distance is effectively obtained.

With regard to a finite collection of piecewise constant ontology score functions $\sum_K = \{f_1, \cdots, f_K\}$ on $V$, with $K \ge 1$. Let $F$ be a collection of ontology score functions. $\overline{f}_K$ is called a median ontology score function for $\sum_K$ with respect to $F$ if

$$\overline{f}_K = \operatorname{argmin}_{f\in F} \Delta_K(f),$$

where $\Delta_K(f) = \sum_{k=1}^{K} d_V(f, f_k)$ for any $f \in F$.

We can get empirical median ontology score function in a similar way, but the true distance $d_V$ should be displaced by its empirical counterpart $d_{\widehat{\tau}_V}$.

From the standard randomization mechanism view, a randomized ontology score function can be regarded as a random element, depending on both a random variable $Z$ and the training sample $S = (S_1, S_2, \cdots, S_k)$. It has the form like $\widehat{f}_n(\cdot, Z)$ which takes values over a measurable space $Z$. The AUC of a randomized ontology score function $\widehat{f}_n(\cdot, Z)$

is given by:

$$\begin{aligned}&\mathrm{AUC}(\widehat{f}_n(\cdot,Z))\\=&\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{\mathbb{P}\{\widehat{f}_n(V,Z)<\widehat{f}_n(V',Z)|(Y_a,Y_b)=(a,b)\}\\&+\frac{1}{2}\mathbb{P}\{\widehat{f}_n(V,Z)=\widehat{f}_n(V',Z)|(Y_a,Y_b)=(a,b)\}\}.\end{aligned}$$

We can say that a randomized ontology score function $\widehat{f}_n$ is AUC-consistent (or, strongly AUC-consistent) if

$$\mathrm{AUC}(\widehat{f}_n(\cdot,Z))\to\mathrm{AUC}^*$$

when $n\to\infty$ holds in probability (or, almost-surely).

Let $K\geq 1$. For fixed $S$, we may draw $K$ Independent and identically distributed copies $Z_1,\cdots,Z_K$ of $Z$, obtaining the collection $\widehat{\sum}_K$ of ontology score functions $\widehat{f}_n(\cdot,Z_j)$, $1\leq j\leq K$. Let $F$ be a collection of ontology score functions and assume that $\widehat{f}_K$ is a median ontology score function for the profile $\widehat{\sum}_K$ with respect to $F$ in the sense of AUC-consistency definition. The result stated as follows implies that AUC-consistency is guarded for a median ontology score function of AUC-consistent randomized ontology score functions.

*Theorem 1:* Let $K\geq 1$ and $F$ be a class of ontology score functions. Suppose that
(1) It satisfies the assumptions on the distribution of $(V,Y)$ in Lemma 2.
(2) The randomized ontology score function $\widehat{f}_n(\cdot,Z_j)$ is AUC-consistent (or, strongly AUC-consistent).
(3) For all $n$, $K\geq 1$ and any sample $S=(S_1,S_2,\cdots,S_k)$, there exists a median ontology score function rule $\widehat{f}_K\in F$ for the collection $\{\widehat{f}_n(\cdot,Z_j),1\leq j\leq K\}$ with respect to $F$.
(4) $F*\cap F\neq\emptyset$.

Then, the ontology score function $\widehat{f}_K$ is AUC-consistent (or, strongly AUC-consistent).

**Proof.** By virtue of Lemma 1, we infer

$$\mathrm{AUC}^*-\mathrm{AUC}(\overline{f}_K)\leq\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{d_V(f^*,\overline{f}_K)}{2p_a^{a,b}p_b^{a,b}},$$

for each $f^*\in F^*$. Apply triangular inequality to the distance $d_V$ between preorders on $V$, we obtain

$$d_V(f^*,\overline{f}_K)\leq d_V(f^*,\widehat{f}_n(\cdot,Z_j))+d_V(\widehat{f}_n(\cdot,Z_j),\overline{f}_K),$$

for each $j\in\{1,\cdots,K\}$. Averaging then over $j$ and according to the fact that

$$\sum_{j=1}^{k}d_V(\widehat{f}_n(\cdot,Z_j),\overline{f}_K)\leq\sum_{j=1}^{K}d_V(\widehat{f}_n(\cdot,Z_j),f^*),$$

if $f^*\in F$, we have

$$d_V(f^*,\overline{f}_K)\leq\frac{2}{K}\sum_{j=1}^{K}d_V(\widehat{f}_n(\cdot,Z_j),f^*).$$

The desired conclusion is finally drawn from the consistency assumption of the randomized ontology score function and the fact described in Lemma 2. □

If we allow $F$ to depend on sample cardinality $n$ and only suppose the existence of $\overline{f}_n^*\in F_n$ satisfies $\mathrm{AUC}(\overline{f}_n^*)\to\mathrm{AUC}^*$ as $n\to\infty$ (omit the condition $F*\cap F\neq\emptyset$), the above statement becomes

$$\begin{aligned}&\mathrm{AUC}^*-\mathrm{AUC}(\overline{f}_K)\\\leq&\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{2p_a^{a,b}p_b^{a,b}}\{\frac{2}{K}\sum_{j=1}^{K}d_V(\widehat{f}_n(\cdot,Z_j),f^*)\\&+d_V(\tilde{f}_n^*,f^*)\},\end{aligned}$$

which implies that the AUC consistency still holds true for the median.

One important fact we should highlight here is that the last assumption in the theory stated above implies that the class $F$ of candidate median ontology score functions includes at least one optimal ontology score function, which can be discarded at the cost of an extra bias term in the rate bound. Then, the consistency results can be attained by picking the median ontology score function, for every $n$, in a class $F_n$ such that there exists a sequence $\tilde{f}_n\in F_n$ satisfies $\mathrm{AUC}(\tilde{f}_n)\to\mathrm{AUC}^*$ as $n\to\infty$. This note contains the special case where $\widehat{f}_n(\cdot,Z_j)$ is a piecewise constant ontology score function with a range of cardinality $r_n\uparrow\infty$ and the median is taken over the set $F_n$ of ontology score functions where its cardinal number not larger than $r_n^K$. The bias term, under mild smoothness conditions on ROC, is then of order $O(\frac{1}{r_n^K})$.

Median calculation is heavily relied on empirical versions of the probabilistic Kendall $\tau$, in the practical sense. The result stated as follows implies that the existence of asymptotically median ontology score functions with respect to $d_V$, provided that the class $F$ over the median is calculated with low complexity. Formally, we formulate the result in view of a VC major class of ontology score functions with finite dimension. For each $f\in F$, we denote

$$\widehat{\Delta}_{K,m}(f)=\sum_{j=1}^{K}\widehat{d}_V(f,f_j),$$

where $\widehat{d}_V$ is the predictor of $d_V$, which is based on $m\geq 1$ independent copies of $V$.

*Theorem 2:* For given $K\geq 1$. Let $\sum_K=\{f_1,\cdots,f_K\}$ be a finite collection ontology score functions and $F$ be a class of ontology score functions which is a VC major class. Considering the empirical median ontology score function $\tilde{f}_m=\mathrm{argmin}_{f\in F}\Delta_K(f)$. Then, as $m\to\infty$, we get

$$\Delta_K(\tilde{f}_m)\to\min_{f\in F}\Delta_K(f)$$

holding for probability 1.

**Proof.** Clearly,

$$\begin{aligned}&\Delta_K(\tilde{f}_m)-\min_{f\in F}\Delta_K(f)\\\leq&2\sup_{f\in F}|\widehat{\Delta}_{K,m}(f)-\Delta_K(f)|\\\leq&2\sum_{j=1}^{K}\sup_{f\in F}|\widehat{d}_V(f,f_j)-d_V(f,f_j)|.\end{aligned}$$

Now, it follows from the strong law of large numbers for $U$-processes such that

$$\sup_{f\in F}|\widehat{d}_V(f,f_j)-d_V(f,f_j)|\to 0$$

as $N \to \infty$, for each $j = 1, \cdots, K$. In terms of central limit theorem for $U$-processes, we get the convergence rate: $O(m^{-\frac{1}{2}})$. □

In next result, we consider that the empirical ontology score function depends heavily on two data samples. The training sample $S$, completed by the randomization on $Z$, results in a collection of ontology score functions which are instances of the randomized ontology score function. Then a sample $S'(m) = \{S_1^{'}, \cdots, S_k^{'}\}$ is used to calculate the empirical median, where $S_a^{'} = m_a$ for $a = 1, \cdots, k$, and $m = \sum_{a=1}^{k} m_a$. Combining with the two preceding results, we finally derive the consistency conclusion from the ontology score function.

*Corollary 1:* For given $K \geq 1$. $F$ is a VC major class of ontology score function. Considering a training sample $S$ of size $n$ with independent and identically distributed in copies of $(V, Y)$ and a sample $S'(m)$ of size $m$ with independent and identically distributed in copies of $V$. $\widehat{\sum}_K$ is collection of randomized ontology score functions $\widehat{f}_n(\cdot, Z_j)$ in $F$ establish out of $S$. The empirical median of $\widehat{\sum}_K$ with respect to $F$ attained from the sample set $S'(m)$ is denoted by $\widehat{f}_{n,m}$. If the assumptions of Theorem 1 are fully true, then we get

$$\mathrm{AUC}(\widehat{f}_{n,m}) \xrightarrow{\mathbb{P}} \mathrm{AUC}^*$$

as $n, m \to \infty$.

**Proof.** Substitute the statement for Theorem 1, we obtain

$$\begin{aligned} d_V(f^*, \widehat{f}_{n,m}) \quad \leq \quad & \frac{1}{K}\sum_{j=1}^{K} d_V(\widehat{f}_n(\cdot, Z_j), f^*) \\ & + \frac{1}{K}\sum_{j=1}^{K} d_V(\widehat{f}_n(\cdot, Z_j), \widehat{f}_{n,m}). \end{aligned}$$

Using the technologies similarity as in Theorem 2's proof, we get

$$\begin{aligned} & \frac{1}{K}\sum_{j=1}^{K}\{d_V(\widehat{f}_n(\cdot, Z_j), \widehat{f}_{n,m}) - d_V(\widehat{f}_n(\cdot, Z_j), \overline{f}_K)\} \\ \leq \quad & 2 \sup_{(f,f')\in F^2} |\widehat{d}_V(f, f') - d_V(f, f')|. \end{aligned}$$

Applying strong Law of Large Numbers for $U$-processes again, we obtain that the term on the right hand side of the bound above tends to 0 as $m \to \infty$. Thus, the desired conclusion is immediately derived from Theorem 1. □

The conclusions described above can be extended to any median ontology score function relying on a pseudo-metric $d$ on the collection of preorders on $F$ which is equivalent to $d_V$, i.e., $c_1 d_V \leq d \leq c_2 d_V$ with $0 < c_1 \leq c_2 < \infty$.

## V. Conclusion

In standard multi-dividing ontology algorithm, the quality of optimal ontology score function heavily depends on the choice of the loss function, and the characteristics of its algorithm are often influenced by various mathematical properties of the loss function. Moreover, standard multi-dividing ontology algorithm has special requirement for the ontology graph structure, which greatly limits the application of the algorithm. In this paper, we propose the new standard of optimal ontology function for multi-dividing method, and such standard is associated with a vertex splitting strategy, which is closely linked with the Kendall $\tau$. The new multi-dividing ontology algorithm theoretically gets rid of the choice of loss function, and relaxes the ontology graph structure. Finally, the convergence for new ontology multi-dividing algorithm has been proved.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References


[1] J. P. Qiu and W. Lou, “Constructing An Information Science Resource Ontology Based on the Chinese Social Science Citation Index,” ASLIB Journal of Information Management, bol. 66, no. 2, pp. 202-218, 2014.

[2] O. Arsene, I. Dumitrache, and I. Mihu, “Medicine Expert System Dynamic Bayesian Network and Ontology Based,” Expert Systems with Applications, vol. 38, no. 12, pp. 15253-15261, 2011.

[3] A. Mehler, O. Pustylnikov, and N. Diewald, “Geography of Social Ontologies: Testing a Variant of the Sapir-Whorf Hypothesis in the Context of Wikipedia,” Computer Speech and Language, vol. 25, no. 3, pp. 716-740, 2011.

[4] X. Su and J. Gulla, “Semantic Enrichment for Ontology Mapping,” The 9th International Conference on Information Systems (NLDB), 2004, pp. 217–228.

[5] L. L. Zhu, X. Z. Min, and W. Gao, “Algorithm of Ontology Similarity Measure Based on Similarity Kernel Learning,” International Journal of Computers & Technology, vol. 14, no. 12, pp. 6304-6309, 2015.

[6] L. L. Zhu and W. Gao, “Ontology Similarity Measuring and Ontology Mapping Based on New Optimization Model in Multi-Dividing Setting,” Journal of Computational Information Systems, vol. 11, no. 1, pp. 377-386, 2015.

[7] W. Gao, L. L. Zhu, and Y. Guo, “Multi-Dividing Infinite Push Ontology Algorithm,” Engineering Letters, vol. 23, no. 3, pp. 132-139, 2015.

[8] W. Gao, J. Z. Wu, and L. L. Zhu, “Ontology Optimization Strategies for Sparse Vector Learning Using Gradient Descent Tricks,” Journal of Computational Information Systems, vol 11, no. 17, pp. 6393-6402, 2015.

[9] Y. Gao and W. Gao, “Ontology Sparse Vector Learning Based on Accelerated First-Order Method,” The Open Cybernetics & Systemics Journal, vol. 9, pp. 657-662, 2015.

[10] M. H. Lan, Y. J. Ren, J. Xu, and W. Gao, “Ontology Similarity Computation Using $k$-Partite Ranking Method,” Journal of Computer Applications, vol. 32, no. 4, pp. 1094–1096, 2012.

[11] W. Gao and L. Shi, “Szeged Related Indices of Unilateral Polyomino Chain and Unilateral Hexagonal Chain,” IAENG International Journal of Applied Mathematics, vol. 45, no. 2, pp. 138-150, 2015.

[12] M. H. Lan, J. Xu, and W. Gao “Ontology Feature Extraction via Vector Learning Algorithm and Applied to Similarity Measuring and Ontology Mapping,” IAENG International Journal of Computer Science, vol. 43, no. 1, pp. 10-19, 2016.

[13] W. Gao and T. W. Xu, “Stability Analysis for Ontology Similarity Computation,” Abstract and Applied Analysis, Vol. 2013, 2013, Article ID 174802, 9 pages, http://dx.doi.org/10.1155/2013/174802.

[14] W. Gao, Y. Gao, and Y. G. Zhang, “Strong and Weak Stability of $k$-Partite Ranking Algorithm,” Journal of Information, vol. 15, no. 11A, pp. 4585–4590, 2012.

[15] W. Gao and T. W. Xu, “Characteristics of Optimal Function for Ontology Similarity Measure via Multi-Dividing,” Journal of networks, vol. 7, no. 8, pp. 1251–1259, 2012.

[16] W. Gao and L. L. Zhu, “Gradient Learning Algorithms for Ontology Computing,” Computational Intelligence and Neuroscience, Volume 2014, Article ID 438291, 12 pages, http://dx.doi.org/10.1155/2014/438291.

[17] 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.

[18] W. Gao, T. W. Xu, J. H. Gan, J. X. Zhou, “Linear Statistical Analysis of Multi-Dividing Ontology Algorithm,” Journal of Information and Computational Science, vol. 11, no. 1, pp. 151–159, 2014.

[19] E. DeLong, D. DeLong, and D. Clark, “Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach,” Biometrics, vol. 44, pp. 837–845, 1988.

[20] J. Hanley and B. McNeil, “The Meaning and Use of The Area Under A Receiver Operating Characteristic (ROC) Curve,” Radiology, vol. 143, pp. 29–36, 1982.
[21] S. Agarwal, T. Graepel, R. Herbrich, S. Har-Peled, and D. Roth, “Generalization Bounds for The Area Under the ROC Curve,” Journal of Machine Learning Research, vol 6, pp. 393–425, 2005.
[22] J. Barthelemy and B. Montjardet, “The Median Procedure in Cluster Analysis and Social Choice Theory,” Mathematical Social Sciences, vol. 1, pp. 235–267, 1981.

**L** inli Zhu, male, was born in the city of Badong, Hubei Province, China on Sep. 20, 1975. He got Master degrees on computer software and theory from Yunnan normal university in 2007. Now, he acts as associate professor in the department of computer engineering, Jiangsu University of Technology. As a researcher in computer science, his interests are covered two disciplines: computer network and artificial intelligence.

Weige Tao received the B.E. degree from Anhui Normal University, China in 1997, the MA.Eng degree from Nanjing University of Science and Technology, China in 2011. Since 2013, he is an associate professor in the School of Electricity and Information Engineering, Jiangsu University of Technology, China. His current research interests include wireless sensor networks, short-range wireless communications, network convergence, and Beidou navigation and positioning.

Xiaozhong Min received the B.E. degree from JiangsSu Normal University, China in 1997, the MA.Eng degree from Soochow University, China in 2002. Since 2005, he is a IT Engineer in the School of Information Center, Jiangsu University of Technology, China. His current research interests include software engineer, component development, database study.

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

He acted as lecturer in the department of information, Yunnan Normal University from July 2012 to December 2015. Now, he acts as associate professor in the department of information, Yunnan Normal University. As a researcher in computer science and mathematics, his interests are covering two disciplines: Graph theory, Statistical learning theory, Information retrieval, and Artificial Intelligence.