计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200100129

决定图框架下本体学习算法的稳定性分析

朱林立[1,2]　华　钢[2]　高　炜[3]
1 江苏理工学院计算机工程学院　江苏 常州 213001
2 中国矿业大学信息与控制工程学院　江苏 徐州 221116
3 云南师范大学信息学院　昆明 650500

摘　要　传统的本体算法采用启发式的方法来计算语义相似度,而随着本体处理数据量的日益增大,越来越多的机器学习方法被用于本体函数的获取。稳定性是本体学习算法的必要条件,它要求在本体样本集做轻微改动的情况下不会对得到的最优本体函数产生本质的改变。文中研究了在本体样本集的依赖关系由图结构决定的框架下,本体学习算法的稳定性和对应的统计学特征。首先对传统的PO和LTO一致稳定条件进行分析;其次在大样本情况下扩展一致稳定性条件,提出Pk和LkO一致稳定性并得到相关的理论结果;最后把替换本体样本和删除本体样本两种样本进行变换组合,提出在大本体样本前提下的组合一致稳定性概念,并利用统计学习理论的方法得到一般结果。此外,在各类稳定性条件下,对满足m-独立条件的本体学习算法的广义界进行了讨论。

关键词:本体;机器学习;稳定性;样本容量;广义界
中图法分类号　TP391

Stability Analysis of Ontology Learning Algorithm in Decision Graph Setting

ZHU Lin-li[1,2],HUA Gang[2] and GAO Wei[3]
1 School of Computer Engineering,Jiangsu University of Technology,Changzhou,Jiangsu 213001,China
2 School of Information and Control Engineering,China University of Mining and Technology,Xuzhou,Jiangsu 221116,China
3 School of Information,Yunnan Normal University,Kunming 650500,China

Abstract　Traditional ontology algorithms use heuristic tricks to calculate semantic similarity. With the increasing amount of data processed by ontology,more and more machine learning technologies are applied to get ontology functions. Stability is a necessary condition for ontology learning algorithms which requires that there is no substantial influence on the obtained optimal ontology function if the ontology sample set is slightly changed. This paper studies the stability and corresponding statistical characteristics of ontology learning algorithms in the setting that the dependency relation of ontology samples are characterized by a graph. Firstly,the traditional PO and LTO uniform stability conditions are analyzed. Then,the extended uniform stability conditions Pk and LkO for large samples are proposed,and related theoretical results are obtained. Finally,two sample transformations (replacement ontology samples and delete ontology samples) are combined to bring forward the concept of combined uniform stability in setting of large ontology samples,and general results are yielded by using statistical learning theory. In addition,under various stability conditions,the generalized bounds of ontology learning algorithms that satisfy the m-independent condition are discussed.

Keywords　Ontology,Machine learning,Stability,Sample capacity,Generalized bound

1　引言

本体作为一种结构化存储工具,已被广泛应用于医学[1-3]、生物学[4-8]、地理学[9-10]等多个学科。近年来,相关学者尝试将各种机器学习方法和本体相融合,通过本体样本的学习来获取本体函数,进而对本体概念之间的相似度进行计算。这方面的相关研究可参考文献[11-14]。

在本体学习算法中,最终得到的本体函数是通过本体样本学习得到的,换句话说,在一定程度上样本的好坏可以影响最后的结果。从直觉上来讲,设计的本体学习算法不能因为本体结构的稍加改变而受到很大的影响,如随着时间的推移,有更多的相关概念加入这个领域本体,从而使得本体的概念增多,结构也相应地发生变化。如果将新增加的本体图顶点加入本体样本中参与学习,那么得到的本体函数与之前的相比应该不会发生本质性的变化。因此,衍生出本体算法稳定性的概念,即增加、减少或更换少量的本体样本,最后得到的

到稿日期:2020-01-19　返修日期:2020-03-20　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金项目(51574232)
This work was supported by the National Natural Science Foundation of China (51574232).
通信作者:朱林立(zhulinli@jsut.edu.cn)

最优本体函数不会发生太大的变化。对于一个有效的本体学习算法,稳定性是必要条件,是算法可以被成功执行的前提。这与学习算法泛化性的要求一致,只有满足稳定性的本体学习算法,才能对除样本点之外的同类本体数据发挥作用,否则学习是无效的。

早期有少量论文对本体算法的稳定性进行了初步的讨论[15-16],但它们的共同缺陷是要求本体数据服从独立同分布原则。其基本思路是:定义稳定性规则、研究算法的稳定条件、利用稳定性给出广义界。其中,广义界的证明一般是利用McDiarmid不等式。

近年来,随着研究的推进,在非独立同分布条件下(non-IID)的情况有所进展,特别是假定样本数据分布可以用一个图来表示的特殊情况,这种表示数据依赖关系的图称为决定图。下面将首先介绍符号、变量和一些已知的定义和定理;然后根据LTO和RO稳定性定义给出对应的结论及证明;同时将这个结果推广到大样本容量下,给出替换k个本体样本或者删除k个本体样本点的情景;最后给出组合稳定性定义,并在此基础上讨论本体学习算法的广义界。

2 稳定性和McDiarmid不等式

设n是正整数,且令$[n]=\{1,2,\cdots,n\}$。对于每个$i\in[n]$,设Ω_i为波利希空间(Polish space),$\Omega=\prod_{i\in[m]}\Omega_i$为它们的乘积空间。$\mathbb{N}_+$和$\mathbb{R}_+$分别表示非负整数和非负实数。给定向量$\boldsymbol{c}=(c_1,\cdots,c_n)\in\mathbb{R}_+^n$,函数$f:\Omega\rightarrow\mathbb{R}$称为满足$\boldsymbol{c}$-利普西茨条件,如果对于任意的$\boldsymbol{x}=(x_1,\cdots,x_n)$,$\boldsymbol{x}'=(x_1',\cdots,x_n')\in\Omega$,有$|f(\boldsymbol{x})-f(\boldsymbol{x}')|\leqslant\sum_{i=1}^{n}c_iI(x_i\neq x'_i)$,其中,$I(\cdot)$是真值函数,$c_i$称为函数$f$的第$i$个利普西茨系数。

文献[17-19]考虑的随机向量的决定图是无圈图的情况,即树或森林。无向图G称为随机向量$\boldsymbol{X}=(X_1,\cdots,X_n)$的决定图,如果下列两个条件成立:

1)$V(G)=[n]$;

2)若$I,J\subset[n]$在G中不相邻;

则$\{X_i\}_{i\in I}$和$\{Y_j\}_{j\in J}$是独立的。

给定一个图G和一个森林F,映射$\phi:V(G)\rightarrow V(F)$。若对每个$(u,v)\in E(G)$有$\phi(u)=\phi(v)$或者$(\phi(u),\phi(v))\in E(F)$,则称$(\phi,F)$为图$G$的森林逼近,并记$\Phi(G)$为图$G$的森林逼近的集合。

给定图G和任意一个森林逼近$(\phi,F)\in\Phi(G)$,其中$F=\{T_i\}_{i\in[k]}$。设:

$$\lambda_{(\phi,F)}=\sum_{(u,v)\in E(F)}(|\phi^{-1}(u)|+|\phi^{-1}(v)|)^2+\sum_{i=1}^{k}\min_{u\in V(T_i)}|\phi^{-1}(u)|^2$$

称$\Lambda(G)=\min_{(\phi,F)\in\Phi(G)}\lambda_{(\phi,F)}$为图$G$的森林复杂度。

有了上述定义后,在一般决定图的框架下,McDiarmid不等式有如下形式。

引理1[19] 设函数$f:\Omega\rightarrow\mathbb{R}$满足$\boldsymbol{c}$-利普西茨条件,随机向量$\boldsymbol{X}$的取值属于$\Omega$,$G$是$\boldsymbol{X}$的决定图。则对于任意$t>0$,有下面不等式成立:

$$\mathrm{P}(f(\boldsymbol{X})-\mathrm{E}[f(\boldsymbol{X})]\geqslant t)\leqslant\exp\left(-\frac{2t^2}{\Lambda(G)\|\boldsymbol{c}\|_\infty^2}\right)$$

下面回到本体框架中,设$S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$是容量为$n$的本体样本,其中$V$是实例空间(输入空间),$Y$是输出空间,设$D$是$V\times Y$上的分布。假设所有的本体训练数据点$(v_i,y_i)$有相同的边际分布,且设$G$是本体样本集$S$的决定图。设$l:Y\times f(V)\rightarrow\mathbb{R}_+\cup\{0\}$是本体亏损函数,则关于本体函数$f$和本体亏损函数$l$的经验误差可以表示为:

$$\hat{R}_S(f)=\frac{1}{n}\sum_{i=1}^{n}l(y_i,f(v_i))$$

对应的期望误差为:

$$R(f)=\mathrm{E}_{(v,y)\sim D}[l(y,f(v))]$$

本体学习算法的稳定性要求对样本集S做少量变动之后,不会大范围影响算法的性能。以下是两种常见的小规模样本调整:

(1)将样本集S中的第$i(i\in\{1,\cdots,n\})$个样本(v_i,y_i)替换为(v_i',y_i')。对应的样本集变为$S^i=\{(v_1,y_1),\cdots,(v_{i-1},y_{i-1}),(v_i',y_i'),(v_{i+1},y_{i+1}),\cdots,(v_n,y_n)\}$。此类变换简称为PO (Replace One)。

(2)将样本集S中的第$i(i\in\{1,\cdots,n\})$个样本(v_i,y_i)和第$j(j\in\{1,\cdots,n\},i\neq j)$个样本$(v_j,y_j)$删除。对应的样本集变为:$S^{\backslash i,j}=\{(v_1,y_1),\cdots,(v_{i-1},y_{i-1}),(v_{i+1},y_{i+1}),\cdots,(v_{j-1},y_{j-1}),(v_{j+1},y_{j+1}),\cdots,(v_n,y_n)\}$。此类变换简称为LTO (leave-Two-Out)。

对于本体学习算法A,设$f_S^A:V\rightarrow Y$为假设函数,即本体算法A从样本S中学习得到。上述两种样本变换对应下面两种一致稳定性假设。

定义1 给定正整数n,本体学习算法A称为关于本体亏损函数l的$\beta_1(n)$-一致稳定。若对于任意$i\in\{1,\cdots,n\}$,$S\in(V\times Y)^n$,$(v_i',y_i')\in V\times Y$,则有:

$$|l(y,f_S^A(v))-l(y,f_{S^i}^A(v))|\leqslant\beta_1(n)$$

定义2 给定正整数n,本体学习算法A称为关于本体亏损函数l的$\beta_2(n)$-一致稳定。若对于任意$i,j\in\{1,\cdots,n\}$ $(i\neq j)$,$S\in(V\times Y)^n$,则有:$|l(y,f_S^A(v))-l(y,f_{S^{\backslash i,j}}^A(v))|\leqslant\beta_2(n)$。

设$\Phi_A(S)=R(f_S^A)-\hat{R}_S(f_S^A)$,该映射$\Phi_A:(V\times Y)^n\rightarrow\mathbb{R}$在下面的计算中发挥了关键性作用。

3 PO稳定条件下的广义界分析

下面的结论说明,在PO稳定条件下,$\Phi_A(S)$的期望值偏差可以大概率地被限定在一定的范围内。

引理2 给定n个本体顶点的本体样本$S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$,$G$是$S$的决定图。假设本体学习算法$A$是$\beta_1(n)$稳定的,且本体亏损函数的上界为$M$,则对于任意$t>0$,有:

$$\mathrm{P}(\Phi_A(S)-\mathrm{E}[\Phi_A(S)]\geqslant t)\leqslant\exp\left(-\frac{2nt^2}{\Lambda(G)(2n\beta_1(n)+M)^2}\right)$$

证明:首先需要证明,若本体学习算法A满足$\beta_1(n)$稳定性,则对任意$(V\times Y)^n$中的样本$S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$和$S^i=\{(v_1,y_1),\cdots,(v_{i-1},y_{i-1}),(v_i',y_i'),(v_{i+1},y_{i+1}),\cdots,(v_n,y_n)\}$(其中$(v_i',y_i')$的边际分布也是$D$),有$|\Phi_A(S)-\Phi_A(S^i)|\leqslant2\beta_1(n)+\frac{M}{n}$成立。

首先注意到:

$$|R(f_S^A)-R(f_{S^i}^A)|$$
$$=|E_D[l(y,f_S^A(x))]-E_D[l(y,f_{S^i}^A(x))]|$$
$$=|E_D[l(y,f_S^A(x))-l(y,f_{S^i}^A(x))]|$$
$$\leqslant\beta_1(n)$$

其次,有:

$$|\hat{R}_S(f_S^A)-\hat{R}_{S^i}(f_{S^i}^A)|$$
$$=|\frac{1}{n}\sum_{(v_j,y_j)\in S}l(y_j,f_S^A(v_j))-\frac{1}{n}\sum_{(v_j,y_j)\in S^i}l(y_j,f_{S^i}^A(v_j))|$$
$$=\frac{1}{n}|\sum_{(v_j,y_j)\in S}l(y_j,f_S^A(v_j))-\sum_{(v_j,y_j)\in S^i}l(y_j,f_{S^i}^A(v_j))|$$
$$\leqslant\frac{1}{n}(\sum_{j\neq i}|l(y_j,f_S^A(v_j))-l(y_j,f_{S^i}^A(v_j))|+|l(y_i,f_S^A(v_i))-l(y_i',f_{S^i}^A(v_i'))|)$$
$$\leqslant\beta_1(n)+\frac{M}{n}$$

进而,有:

$$|\Phi_A(S)-\Phi_A(S^i)|$$
$$=|(R(f_S^A)-\hat{R}_S(f_S^A))-(R(f_{S^i}^A)-\hat{R}_{S^i}(f_{S^i}^A))|\leqslant$$
$$|R(f_S^A)-R(f_{S^i}^A)|+|\hat{R}_S(f_S^A)-\hat{R}_{S^i}(f_{S^i}^A)|$$
$$\leqslant2\beta_1(n)+\frac{M}{n}$$

将其与前文的引理1相结合,即得引理2。

引理3刻画了$\Phi_A(S)$的期望值的上界。

引理3 给定n个本体顶点的本体样本$S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$,$G$是$S$的决定图。假设本体学习算法$A$是$\beta_1(i)$稳定的($i\leqslant n$是任意的),且图$G$的最大度为$\Delta$。设$\beta_1(n,\Delta)=\max_{i\in[0,\Delta]}\beta_1(n-i)$,有:

$$E[\Phi_A(S)]\leqslant2(\Delta+1)\beta_1(n,\Delta)$$

证明:需要证明在引理3条件成立的情况下,有:

$$\max_{(v_j,y_j)\in S}E_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]\leqslant2(\Delta+1)\beta_1(n,\Delta)$$

对于任意$i\in\{1,\cdots,n\}$,设$N_G^+(i)=\{j_1,\cdots,j_{n_i}\}$满足$j_{k-1}>j_k$。对于任意$k\in\{1,\cdots,n_i\}$,定义$S^{(i,0)}=S$,且$S^{(i,k)}$为在$S^{(i,k-1)}$中替换第$j_k$个本体样本后得到。根据本体算法$A$的稳定性假设,有:

$$|l(y,f_{S^{(i,k-1)}}^A(v))-l(y,f_{S^{(i,k)}}^A(v))|\leqslant\beta_1(n,\Delta)$$

同时注意到:

$$l(y,f_S^A(v))=\sum_{k=1}^{n_i}(l(y,f_{S^{(i,k-1)}}^A(v))-l(y,f_{S^{(i,k)}}^A(v)))+l(y,f_{S^{(i,n_i)}}^A(v))$$

$$l(y_i,f_S^A(v_i))=\sum_{k=1}^{n_i}(l(y_i,f_{S^{(i,k-1)}}^A(v_i))-l(y_i,f_{S^{(i,k)}}^A(v_i)))+l(y_i,f_{S^{(i,n_i)}}^A(v_i))$$

因此,得到:

$$l(y,f_S^A(v))-l(y_i,f_S^A(v_i))$$
$$=\sum_{k=1}^{n_i}(l(y,f_{S^{(i,k-1)}}^A(v))-l(y,f_{S^{(i,k)}}^A(v)))-\sum_{k=1}^{n_i}(l(y_i,f_{S^{(i,k-1)}}^A(v_i))-l(y_i,f_{S^{(i,k)}}^A(v_i)))+l(y,f_{S^{(i,n_i)}}^A(v))-l(y_i,f_{S^{(i,n_i)}}^A(v_i))$$
$$\leqslant\sum_{k=1}^{n_i}|l(y,f_{S^{(i,k-1)}}^A(v))-l(y,f_{S^{(i,k)}}^A(v))|+\sum_{k=1}^{n_i}|l(y_i,f_{S^{(i,k-1)}}^A(v_i))-l(y_i,f_{S^{(i,k)}}^A(v_i))|+l(y,f_{S^{(i,n_i)}}^A(v))-l(y_i,f_{S^{(i,n_i)}}^A(v_i))$$
$$\leqslant2n_i\beta_1(n,\Delta)+l(y,f_{S^{(i,n_i)}}^A(v))-l(y_i,f_{S^{(i,n_i)}}^A(v_i))$$

由于(v_i,y_i)和(v,y)关于$S^{(i,n_i)}$是独立的,可得:

$$\mathrm{E}_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]$$
$$=\mathrm{E}_{S,(v,y)}[l(y,f_{S^{(i,n_i)}}^A(v))-l(y_i,f_{S^{(i,n_i)}}^A(v_i))]+2n_i\beta_1(n,\Delta)$$
$$\leqslant\mathrm{E}_{S,(v,y)}[l(y,f_{S^{(i,n_i)}}^A(v))]-l(y_i,f_{S^{(i,n_i)}}^A(v_i))]+2(\Delta+1)\beta_1(n,\Delta)$$
$$=\mathrm{E}_{S,(v,y)}[l(y,f_{S^{(i,n_i)}}^A(v))]-\mathrm{E}_S[l(y_i,f_{S^{(i,n_i)}}^A(v_i))]-\mathrm{E}_{S^{(i,n_i)},(v_i,y_i)}[l(y_i,f_{S^{(i,n_i)}}^A(v_i))]$$
$$=2(\Delta+1)\beta_1(n,\Delta)+\mathrm{E}_{S^{(i,n_i)},(v,y)}[l(y,f_{S^{(i,n_i)}}^A(v))]-\mathrm{E}_{S^{(i,n_i)},(v_i,y_i)}[l(y_i,f_{S^{(i,n_i)}}^A(v_i))]$$
$$=2(\Delta+1)\beta_1(n,\Delta)$$

最后,有:

$$\mathrm{E}[\Phi_A(S)]=\mathrm{E}_S[\mathrm{E}_{(v,y)}[l(y,f_S^A(v))]-\frac{1}{n}\sum_{i=1}^{n}l(y_i,f_S^A(v_i))]$$
$$=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]$$
$$\leqslant2(\Delta+1)\beta_1(n,\Delta)$$

引理3证毕。

结合引理2和引理3,可得到定理1。

定理1 给定n个本体顶点的本体样本$S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$,$G$是$S$的决定图。假设本体学习算法$A$是$\beta_1(i)$稳定的($i\leqslant n$是任意的),本体亏损函数的上界为$M$,且图$G$的最大度为$\Delta$。设$\beta_1(n,\Delta)=\max_{i\in[0,\Delta]}\beta_1(n-i)$,则对任意$\delta\in(0,1)$,下式以置信度$1-\delta$成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(\Delta+1)\beta_1(n,\Delta)+\frac{2n\beta_1(n)+M}{n}\sqrt{\frac{\Lambda(G)\ln(\frac{1}{\delta})}{2}}$$

设$m,n\in\mathbb{N}_+$,随机变量序列$\{X_i\}_{i=1}^n$称为m-独立,如果对任意$i\in[n-m-1]$,$\{X_j\}_{j=1}^i$独立于$\{X_j\}_{j=i+m+1}^n$。若本体样本集满足m-独立的特殊条件,则有定理2。

定理2 给定n个本体顶点的m-独立本体样本$S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$,$G$是$S$的决定图。假设本体学习算法$A$是$\beta_1(i)$稳定的($i\leqslant n$是任意的),本体亏损函数的上界为$M$。则对任意$\delta\in(0,1)$,下式以置信度$1-\delta$成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(m+1)\beta_1(n,2m)+(2n\beta_1(n)+M)\sqrt{\frac{2m\ln(\frac{1}{\delta})}{n}}$$

4 LTO稳定条件下的广义界分析

利用类似的方法,可以得到在本体学习算法满足LTO稳定的条件下算法的统计学特征。引理4说明在LTO稳定的条件下,$\Phi_A(S)$的期望值偏差可以大概率地被限定在一定的范围内。

引理4 给定n个本体顶点的本体样本$S=\{(v_1,y_1),\cdots,$

$(v_n,y_n)\}$，G 是 S 的决定图。假设本体学习算法 A 是 $\beta_2(n)$ 稳定的，且本体亏损函数的上界为 M。则对于任意 $t>0$，有：

$$\mathrm{P}(\Phi_A(S)-\mathrm{E}[\Phi_A(S)]\geqslant t)$$
$$\leqslant \exp\left(-\frac{2nt^2}{\Lambda(G)(4n\beta_2(n)+2M)^2}\right)$$

证明：首先需要证明，若本体学习算法 A 满足 $\beta_2(n)$ 稳定性，则对任意 $(V\times Y)^n$ 中的样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$ 和 $S^{i,j}=\{(v_1,y_1),\cdots,(v_{i-1},y_{i-1}),(v_i',y_i'),(v_{i+1},y_{i+1}),\cdots,(v_{j-1},y_{j-1}),(v'_j,y'_j),(v_{j+1},y_{j+1}),\cdots,(v_n,y_n)\}$（其中，$(v_i',y_i')$ 和 (v'_j,y'_j) 的边际分布也是 D，该样本变换是替换 S 中的两个样本），有 $|\Phi_A(S)-\Phi_A(S^{i,j})|\leqslant 4\beta_2(n)+\frac{2M}{n}$ 成立。

首先，注意到：

$$\begin{aligned}
&|R(f_S^A)-R(f_{S^{i,j}}^A)|\\
&\leqslant |R(f_S^A)-R(f_{S^{\backslash i,j}}^A)|+|R(_{S^{\backslash i,j}}^A)-R(f_{S^{i,j}}^A)|\\
&=|\mathrm{E}_D[l(y,f_S^A(x))]-\mathrm{E}_D[l(y,_{S^{\backslash i,j}}^A(x))]|+\\
&\quad|\mathrm{E}_D[l(y,_{S^{\backslash i,j}}^A(x))]-\mathrm{E}_D[l(y,f_{S^{i,j}}^A(x))]|\\
&=|\mathrm{E}_D[l(y,f_S^A(x))-l(y,_{S^{\backslash i,j}}^A(x))]|+|\mathrm{E}_D[l(y,\\
&\quad f_{S^{\backslash i,j}}^A(x))-l(y,f_{S^{i,j}}^A(x))]|\\
&\leqslant 2\beta_2(n)
\end{aligned}$$

其次，有：

$$\begin{aligned}
&|\hat{R}_S(f_S^A)-\hat{R}_{S^{i,j}}(f_{S^{i,j}}^A)|\\
&=\left|\frac{1}{n}\sum_{(v_k,y_k)\in S}l(y_k,f_S^A(v_k))-\frac{1}{n}\sum_{(v_k,y_k)\in S^{i,j}}l(y_k,\right.\\
&\quad\left.f_{S^{i,j}}^A(v_k))\right|\\
&=\frac{1}{n}\left|\sum_{(v_k,y_k)\in S}l(y_k,f_S^A(v_k))-\sum_{(v_k,y_k)\in S^{i,j}}l(y_k,f_{S^{i,j}}^A(v_k))\right|\\
&\leqslant\frac{1}{n}\Big(\sum_{k\neq i,k\neq j}|l(y_k,f_S^A(v_k))-l(y_k,f_{S^{i,j}}^A(v_k))|+|l(y_i,\\
&\quad f_S^A(v_i))-l(y_i',f_{S^{i,j}}^A(v_i'))|+|l(y_j,f_S^A(v_j))-\\
&\quad l(y_j',f_{S^{i,j}}^A(v'_j))|\Big)\\
&\leqslant\frac{1}{n}\Big(\sum_{k\neq i,k\neq j}|l(y_k,f_S^A(v_k))-l(y_k,f_{S^{\backslash i,j}}^A(v_k))|+\sum_{k\neq i,k\neq j}\\
&\quad|l(y_k,_{S^{\backslash i,j}}^A(v_k))-l(y_k,f_{S^{i,j}}^A(v_k))|+|l(y_i,f_S^A(v_i))-\\
&\quad l(y_i',f_{S^{i,j}}^A(v_i'))|+|l(y_j,f_S^A(v_j))-l(y'_j,\\
&\quad f_{S^{i,j}}^A(v'_j))|\Big)\\
&\leqslant 2\beta_2(n)+\frac{2M}{n}
\end{aligned}$$

进而，有：

$$\begin{aligned}
&|\Phi_A(S)-\Phi_A(S^{i,j})|\\
&=|(R(f_S^A)-\hat{R}_S(f_S^A))-(R(f_{S^{i,j}}^A)-\hat{R}_{S^{i,j}}(f_{S^{i,j}}^A))|\\
&\leqslant|R(f_S^A)-R(f_{S^{i,j}}^A)|+|\hat{R}_S(f_S^A)-\hat{R}_{S^{i,j}}(f_{S^{i,j}}^A)|\\
&\leqslant 4\beta_2(n)+\frac{2M}{n}
\end{aligned}$$

将其与前文的引理 1 相结合，即得引理 4。

引理 5 刻画了在 LTO 框架下 $\Phi_A(S)$ 的期望值的上界。

引理 5 给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\beta_2(i)$ 稳定的（$i\leqslant n$ 是任意的），且图 G 的最大度为 Δ。设 $\beta_2(n,\Delta)=\max\limits_{i\in[0,\Delta]}\beta_2(n-i)$，有：

$$\mathrm{E}[\Phi_A(S)]\leqslant 2(\Delta+1)\beta_2(n,\Delta)$$

证明：需要证明，在引理 5 条件成立的情况下，有：

$$\max_{(v_j,y_j)\in S}\mathrm{E}_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]$$
$$\leqslant 2(\Delta+1)\beta_2(n,\Delta)$$

对于任意 $i\in\{1,\cdots,n\}$，设 $N_G^+(i)=\{j_1,\cdots,j_{n_i}\}$ 满足 $j_{k-1}>j_k$。对任意 $k\in\{1,\cdots,n_i\}$ 定义 $S^{(i,-1)}=S^{(i,0)}=S$，且 $S^{(i,k)}$ 通过在 $S^{(i,k-2)}$ 中删除第 j_k 和 j_{k-1} 两个本体样本后得到。根据本体算法 A 的稳定性假设，有：

$$|l(y,f_{S^{(i,k-2)}}^A(v))-l(y,f_{S^{(i,k)}}^A(v))|\leqslant\beta_2(n,\Delta)$$

同时，注意到：

$$\begin{aligned}
l(y,f_S^A(v))=&\sum_{k=1}^{n_i}(l(y,f_{S^{(i,k-2)}}^A(v))-l(y,f_{S^{(i,k)}}^A(v)))+\\
&l(y,f_{S^{(i,n_i-1)}}^A(v))+l(y,f_{S^{(i,n_i)}}^A(v))
\end{aligned}$$

$$\begin{aligned}
l(y_i,f_S^A(v_i))=&\sum_{k=1}^{n_i}(l(y_i,f_{S^{(i,k-2)}}^A(v_i))-l(y_i,f_{S^{(i,k)}}^A(v_i)))+\\
&l(y_i,f_{S^{(i,n_i-1)}}^A(v_i))+l(y_i,f_{S^{(i,n_i)}}^A(v_i))
\end{aligned}$$

因此，得到：

$$\begin{aligned}
&l(y,f_S^A(v))-l(y_i,f_S^A(v_i))\\
&=\sum_{k=1}^{n_i}(l(y,f_{S^{(i,k-2)}}^A(v))-l(y,f_{S^{(i,k)}}^A(v)))-\sum_{k=1}^{n_i}(l(y_i,\\
&\quad f_{S^{(i,k-2)}}^A(v_i))-l(y_i,f_{S^{(i,k)}}^A(v_i)))+(l(y,f_{S^{(i,n_i)}}^A(v))-\\
&\quad l(y_i,f_{S^{(i,n_i)}}^A(v_i)))+(l(y,f_{S^{(i,n_i-1)}}^A(v))-l(y_i,f_{S^{(i,n_i-1)}}^A\\
&\quad(v_i)))\\
&\leqslant\sum_{k=1}^{n_i}|l(y,f_{S^{(i,k-1)}}^A(v))-l(y,f_{S^{(i,k)}}^A(v))|+\sum_{k=1}^{n_i}|l(y_i,\\
&\quad f_{S^{(i,k-1)}}^A(v_i))-l(y_i,f_{S^{(i,k)}}^A(v_i))|+(l(y,f_{S^{(i,n_i)}}^A(v))\\
&\quad-l(y_i,f_{S^{(i,n_i)}}^A(v_i)))+(l(y,f_{S^{(i,n_i-1)}}^A(v))-l(y_i,\\
&\quad f_{S^{(i,n_i-1)}}^A(v_i)))\\
&\leqslant(l(y,f_{S^{(i,n_i)}}^A(v))-l(y_i,f_{S^{(i,n_i)}}^A(v_i)))+(l(y,f_{S^{(i,n_i-1)}}^A\\
&\quad(v))-l(y_i,f_{S^{(i,n_i-1)}}^A(v_i)))+2n_i\beta_2(n,\Delta)
\end{aligned}$$

由于 (v_i,y_i) 和 (v,y) 关于 $S^{(i,n_i)}$ 和 $S^{(i,n_i-1)}$ 都是独立的，可得：

$$\begin{aligned}
&\mathrm{E}_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]\\
&=\mathrm{E}_{S,(v,y)}[l(y,f_{S^{(i,n_i)}}^A(v))-l(y_i,f_{S^{(i,n_i)}}^A(v_i))+l(y,\\
&\quad f_{S^{(i,n_i-1)}}^A(v))-l(y_i,f_{S^{(i,n_i-1)}}^A(v_i))]+2n_i\beta_2(n,\Delta)\\
&\leqslant\mathrm{E}_{S,(v,y)}[l(y,f_{S^{(i,n_i)}}^A(v))-l(y_i,f_{S^{(i,n_i)}}^A(v_i))+l(y,\\
&\quad f_{S^{(i,n_i-1)}}^A(v))-l(y_i,f_{S^{(i,n_i-1)}}^A(v_i))]+2(\Delta+1)\beta_2(n,\Delta)\\
&=\mathrm{E}_{S,(v,y)}[l(y,f_{S^{(i,n_i)}}^A(v))]-\mathrm{E}_S[l(y_i,f_{S^{(i,n_i)}}^A(v_i))]+\\
&\quad\mathrm{E}_{S,(v,y)}[l(y,f_{S^{(i,n_i-1)}}^A(v))]-\mathrm{E}_S[l(y_i,f_{S^{(i,n_i-1)}}^A(v_i))]+\\
&\quad 2(\Delta+1)\beta_2(n,\Delta)\\
&=\mathrm{E}_{S^{(i,n_i)},(v,y)}[l(y,f_{S^{(i,n_i)}}^A(v))]-\mathrm{E}_{S^{(i,n_i)},(v_i,y_i)}[l(y_i,\\
&\quad f_{S^{(i,n_i)}}^A(v_i))]+\mathrm{E}_{S^{(i,n_i-1)},(v,y)}[l(y,f_{S^{(i,n_i-1)}}^A(v))]-\\
&\quad\mathrm{E}_{S^{(i,n_i-1)},(v_i,y_i)}[l(y_i,f_{S^{(i,n_i-1)}}^A(v_i))]+2(\Delta+1)\beta_2(n,\Delta)\\
&=2(\Delta+1)\beta_2(n,\Delta)
\end{aligned}$$

最后，有：

$$\begin{aligned}
\mathrm{E}[\Phi_A(S)]=&\mathrm{E}_S\Big[E_{(v,y)}[l(y,f_S^A(v))]-\frac{1}{n}\sum_{i=1}^{n}l(y_i,\\
&f_S^A(v_i))\Big]\\
&=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]\\
&\leqslant 2(\Delta+1)\beta_2(n,\Delta)
\end{aligned}$$

引理 5 证毕。

结合引理 4 和引理 5，可得到如下定理 3。

定理 3 给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\beta_2(i)$ 稳定的（$i\leqslant n$ 是任意的），本体亏损函数的上界为 M，且图 G 的最大度为 Δ。设 $\beta_2(n,\Delta)=\max\limits_{i\in[0,\Delta]}\beta_2(n-i)$。则对任意 $\delta\in(0,1)$，下式以置信度 $1-\delta$ 成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(\Delta+1)\beta_2(n,\Delta)+\frac{4n\beta_2(n)+2M}{n}\sqrt{\frac{\Lambda(G)\ln(\frac{1}{\delta})}{2}}$$

在 TWO 框架下，若本体样本集满足 m-独立的特殊条件，则有定理 4。

定理 4 给定 n 个本体顶点的 m-独立本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\beta_2(i)$ 稳定的（$i\leqslant n$ 是任意的），本体亏损函数的上界为 M，则对任意 $\delta\in(0,1)$，下式以置信度 $1-\delta$ 成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(m+1)\beta_2(n,2m)+(4n\beta_2(n)+2M)\sqrt{\frac{2m\ln(\frac{1}{\delta})}{n}}$$

5 替换或删除 k 个本体样本条件下的稳定性和广义界

假设样本容量很大，这时增加替换的样本顶点或删除的本体样本数量，相对于大样本容量而言依然算是样本集的轻微变动。在此假设下，引入下面的一致稳定性定义。

定义 3 设正整数 k,n 满足 $k\ll n$。本体学习算法 A 称为关于本体亏损函数 l 的 $\gamma_k(n)$-一致稳定。若对于任意 $i_1,i_2,\cdots,i_k\in\{1,\cdots,n\}$，$(v'_{i_1},y'_{i_1}),(v'_{i_2},y'_{i_2}),\cdots,(v'_{i_k},y'_{i_k})\in V\times Y$，$S\in(V\times Y)^n$，则有：

$$|l(y,f_S^A(v))-l(y,f_{S^{i_1,i_2,\cdots,i_k}}^A(v))|\leqslant\gamma_k(n)$$

其中，$S^{i_1,i_2,\cdots,i_k}$ 表示在原来本体样本集 S 中，将 (v_{i_1},y_{i_1})，(v_{i_2},y_{i_2})，$\cdots$，(v_{i_k},y_{i_k}) 分别替换为 (v'_{i_1},y'_{i_1})，(v'_{i_2},y'_{i_2})，$\cdots$，(v'_{i_k},y'_{i_k}) 后得到的本体样本集。

定义 4 设正整数 k,n 满足 $k=n$。本体学习算法 A 称为关于本体亏损函数 l 的 $\eta_k(n)$-一致稳定。若对于任意 $i_1,i_2,\cdots,i_k\in\{1,\cdots,n\}$，$S\in(V\times Y)^n$，则有：

$$|l(y,f_S^A(v))-l(y,{}^{A}_{S^{i_1,i_2,\cdots,i_k}}(v))|\leqslant\eta_k(n)$$

其中，$S^{\backslash i_1,i_2,\cdots,i_k}$ 表示在原来本体样本集 S 中将 (v_{i_1},y_{i_1})，(v_{i_2},y_{i_2})，$\cdots$，(v_{i_k},y_{i_k}) 删除后得到的本体样本集。

定义 3 和定义 4 分别称为 Pk 一致稳定性和 LkO 一致稳定性，它们分别刻画了在大本体样本容量条件下，删除部分本体样本，学习算法的稳定性能。当 $k=1$ 时，$\gamma_1(n)$ 即为 $\beta_1(n)$；当 $k=2$ 时，$\eta_2(n)$ 即为 $\beta_2(n)$。

根据之前的证明思路，得到如下扩展的结果。由于证明方法相似，这里不再给出具体的证明。

引理 6 设正整数 k,n 满足 $k\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\gamma_k(n)$ 稳定的，且本体亏损函数的上界为 M，则对于任意 $t>0$，有：

$$P(\Phi_A(S)-E[\Phi_A(S)]\geqslant t)\leqslant\exp(-\frac{2nt^2}{\Lambda(G)(2n\gamma_k(n)+kM)^2})$$

引理 7 设正整数 k,n 满足 $k\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\gamma_k(i)$ 稳定的（$i\leqslant n$ 是任意的），且图 G 的最大度为 Δ。设 $\gamma_k(n,\Delta)=\max\limits_{i\in[0,\Delta]}\gamma_k(n-i)$，有：

$$E[\Phi_A(S)]\leqslant2(\Delta+1)\gamma_k(n,\Delta)$$

定理 5 设正整数 k,n 满足 $k\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\gamma_k(i)$ 稳定的（$i\leqslant n$ 是任意的），本体亏损函数的上界为 M，且图 G 的最大度为 Δ。设 $\gamma_k(n,\Delta)=\max\limits_{i\in[0,\Delta]}\gamma_k(n-i)$。则对任意 $\delta\in(0,1)$，下式以置信度 $1-\delta$ 成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(\Delta+1)\gamma_k(n,\Delta)+\frac{2n\gamma_k(n)+kM}{n}\sqrt{\frac{\Lambda(G)\ln(\frac{1}{\delta})}{2}}$$

定理 6 设正整数 k,n 满足 $k\ll n$。给定 n 个本体顶点的 m-独立本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\gamma_k(i)$ 稳定的（$i\leqslant n$ 是任意的），本体亏损函数的上界为 M，则对任意 $\delta\in(0,1)$，下式以置信度 $1-\delta$ 成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(m+1)\gamma_k(n,2m)+(2n\gamma_k(n)+kM)\sqrt{\frac{2m\ln(\frac{1}{\delta})}{n}}$$

引理 8 设正整数 k,n 满足 $k\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\eta_k(n)$ 稳定的，且本体亏损函数的上界为 M，则对于任意 $t>0$，有：

$$P(\Phi_A(S)-E[\Phi_A(S)]\geqslant t)\leqslant\exp(-\frac{2nt^2}{\Lambda(G)(4n\eta_k(n)+kM)^2})$$

引理 9 设正整数 k,n 满足 $k\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\eta_k(i)$ 稳定的（$i\leqslant n$ 是任意的），且图 G 的最大度为 Δ。设 $\eta_k(n,\Delta)=\max\limits_{i\in[0,\Delta]}\eta_k(n-i)$，有：

$$E[\Phi_A(S)]\leqslant2(\Delta+1)\eta_k(n,\Delta)$$

定理 7 设正整数 k,n 满足 $k\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\eta_k(i)$ 稳定的（$i\leqslant n$ 是任意的），本体亏损函数的上界为 M，且图 G 的最大度为 Δ。设 $\eta_k(n,\Delta)=\max\limits_{i\in[0,\Delta]}\eta_k(n-i)$，则对任意 $\delta\in(0,1)$，下式以置信度 $1-\delta$ 成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(\Delta+1)\eta_k(n,\Delta)+\frac{4n\eta_k(n)+kM}{n}\sqrt{\frac{\Lambda(G)\ln(\frac{1}{\delta})}{2}}$$

定理 8 设正整数 k,n 满足 $k\ll n$。给定 n 个本体顶点的

m-独立本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\eta_k(i)$ 稳定的（$i\leqslant n$ 是任意的），本体亏损函数的上界为 M，则对任意 $\delta\in(0,1)$，下式以置信度 $1-\delta$ 成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(m+1)\eta_k(n,2m)+(4n\eta_k(n)+kM)\sqrt{\frac{2m\ln(\frac{1}{\delta})}{n}}$$

6 组合样本变换下的稳定性和广义界

本节将前面样本变换方式进行组合，即在原有本体样本 S 的条件下，删除 k_1 个样本，再替换剩余 k_2 个样本，其中 $k_1+k_2\ll n$。

定义 5 设非负整数 k_1,k_2,n 满足 $0<k_1+k_2\ll n$。本体学习算法 A 称为关于本体亏损函数 l 的 $\Xi_{k_1,k_2}(n)$-一致稳定。若对于任意 $S\in(V\times Y)^n$，$i_1,i_2,\cdots,i_{k_1}\in\{1,\cdots,n\}$，$j_1,j_2,\cdots,j_{k_2}\in\{1,\cdots,n\}-\{i_1,i_2,\cdots,i_{k_1}\}$，$(v'_{j_1},y'_{j_1})$，$(v'_{j_2},y'_{j_2})$，$\cdots$，$(v'_{j_{k_2}},y'_{j_{k_2}})\in V\times Y$，则有：

$$|l(y,f_S^A(v))-l(y,f^A_{S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}},(v))|$$

$$\leqslant\Xi_{k_1,k_2}(n)$$

其中，$S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}$ 表示在原来本体样本集 S 中删除 $(v_{i_1},y_{i_1}),(v_{i_2},y_{i_2}),\cdots,(v_{i_{k_1}},y_{i_{k_1}})$，并且在剩下的样本集中将 $(v_{j_1},y_{j_1}),(v_{j_2},y_{j_2}),\cdots,(v_{j_{k_2}},y_{j_{k_2}})$ 分别替换为 $(v'_{j_1},y'_{j_1}),(v'_{j_2},y'_{j_2}),\cdots,(v'_{j_{k_2}},y'_{j_{k_2}})$ 后得到的本体样本集。

作为定义 3 和定义 4 的组合，定义 5 考虑了同时删除和替换一小部分本体样本后的稳定性，称 $\Xi_{k_1,k_2}(n)$ 为本体亏损组合一致稳定。当 $k_1=0$ 时，$\Xi_{k_1,k_2}(n)$ 退化为定义 3 中的 Pk 一致稳定性；当 $k_2=0$ 时，$\Xi_{k_1,k_2}(n)$ 退化为定义 4 中的 LkO 一致稳定性。定义 5 的前提是大样本容量，使得 $k_1+k_2\ll n$ 成立。在此假设下，$\Xi_{k_1,k_2}(n)$ 组合稳定性刻画了在本体样本集微小变动下（减少和替换少量本体样本点），算法得到的最优本体函数不会发生本质上的改变。

引理 10 说明在 $\Xi_{k_1,k_2}(n)$ 组合稳定条件下，$\Phi_A(S)$ 的期望值偏差可以大概率地被限定在一定的范围内。

引理 10 设非负整数 k_1,k_2,n 满足 $0<k_1+k_2\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$，$G$ 是 S 的决定图。假设本体学习算法 A 是 $\Xi_{k_1,k_2}(n)$ 稳定的，且本体亏损函数的上界为 M，则对于任意 $t>0$，有：

$$\mathrm{P}(\Phi_A(S)-\mathrm{E}[\Phi_A(S)]\geqslant t)$$

$$\leqslant\exp\left(-\frac{2nt^2}{\Lambda(G)(4n\Xi_{k_1,k_2}(n)+(k_1+k_2)M)^2}\right)$$

证明：首先需要证明，若本体学习算法 A 满足 $\Xi_{k_1,k_2}(n)$ 稳定性，则对任意 $(V\times Y)^n$ 中的样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$ 和 $S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}$（该样本变换是将 S 中的 k_1+k_2 个样本 $(v_{i_1},y_{i_1}),(v_{i_2},y_{i_2}),\cdots,(v_{i_{k_1}},y_{i_{k_1}}),(v_{j_1},y_{j_1}),(v_{j_2},y_{j_2}),\cdots,(v_{j_{k_2}},y_{j_{k_2}})$ 分别替换为 $(v'_{i_1},y'_{i_1}),(v'_{i_2},y'_{i_2}),\cdots,(v'_{i_{k_1}},y'_{i_{k_1}}),(v'_{j_1},y'_{j_1}),(v'_{j_2},y'_{j_2}),\cdots,(v'_{j_{k_2}},y'_{j_{k_2}})$ 后得到的，其中 $(v'_{i_1},y'_{i_1}),(v'_{i_2},y'_{i_2}),\cdots,(v'_{i_{k_1}},y'_{i_{k_1}}),(v'_{j_1},y'_{j_1}),(v'_{j_2},y'_{j_2}),\cdots,(v'_{j_{k_2}},y'_{j_{k_2}})$ 中每个样本的边际分布也是 D），有：

$$|\Phi_A(S)-\Phi_A(S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}})|$$

$$\leqslant 4\Xi_{k_1,k_2}(n)+\frac{(k_1+k_2)M}{n}$$

成立。

首先注意到：

$$\begin{aligned}
&|R(f_S^A)-R(f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}})|\\
&\leqslant|R(f_S^A)-R(f^A_{S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}})|+\\
&\quad|R(f^A_{S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}})-R(f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}})|\\
&=|E_D[l(y,f_S^A(x))]-E_D[l(y,f^A_{S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}}\\
&\quad(x))]|+|E_D[l(y,f^A_{S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}}(x))]-\\
&\quad E_D[l(y,f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(x))]|\\
&=|E_D[l(y,f_S^A(x))-l(y,f^A_{S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}}(x))]|+\\
&\quad|E_D[l(y,f^A_{S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}}(x))-l(y,\\
&\quad f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(x))]|\\
&\leqslant 2\Xi_{k_1,k_2}(n)
\end{aligned}$$

在上式推导过程中，特别要强调的是 $S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}$ 和 $S_S^{A\setminus\{i_1,i_2,\cdots,i_{k_1-1},j_2,\cdots,j_{k_2}\}}$ 这两个样本集合在 $j_1,j_2,\cdots,j_{k_2}$ 位置上替换了不同的本体样本。前者将 $(v_{j_1},y_{j_1}),(v_{j_2},y_{j_2}),\cdots,(v_{j_{k_2}},y_{j_{k_2}})$ 替换为 $(v'_{j_1},y'_{j_1}),(v'_{j_2},y'_{j_2}),\cdots,(v'_{j_{k_2}},y'_{j_{k_2}})$；后者将 $(v_{j_1},y_{j_1}),(v_{j_2},y_{j_2}),\cdots,(v_{j_{k_2}},y_{j_{k_2}})$ 替换为 $(v''_{j_1},y''_{j_1}),(v''_{j_2},y''_{j_2}),\cdots,(v''_{j_{k_2}},y''_{j_{k_2}})$。同时 $(v'_{j_1},y'_{j_1}),(v'_{j_2},y'_{j_2}),\cdots,(v'_{j_{k_2}},y'_{j_{k_2}})$ 和 $(v''_{j_1},y''_{j_1}),(v''_{j_2},y''_{j_2}),\cdots,(v''_{j_{k_2}},y''_{j_{k_2}})$ 中的每个本体样本的边际分布也是 D，且满足 $(v'_{j_1},y'_{j_1})\neq(v''_{j_1},y''_{j_1})$，$(v'_{j_2},y'_{j_2})\neq(v''_{j_2},y''_{j_2})$，$\cdots$，$(v'_{j_{k_2}},y'_{j_{k_2}})\neq(v''_{j_{k_2}},y''_{j_{k_2}})$。这样保证了 $S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}$ 可以通过在 $S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}$ 样本基础之上删除 $i_1,i_2,\cdots,i_{k_1}$ 位置上的本体样本，再替换 $j_1,j_2,\cdots,j_{k_2}$ 位置上的本体样本而得到，进而确保了最后一个不等式成立。

其次，有：

$$\begin{aligned}
&|\hat{R}_S(f_S^A)-\hat{R}_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}})|\\
&=|\frac{1}{n}\sum_{(v_k,y_k)\in S}l(y_k,f_S^A(v_k))-\frac{1}{n}\sum_{(v_k,y_k)\in S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}\\
&\quad l(y_k,f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v_k))|\\
&=\frac{1}{n}|\sum_{(v_k,y_k)\in S}l(y_k,f_S^A(v_k))-\sum_{(v_k,y_k)\in S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}l(y_k,\\
&\quad f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v_k))|\\
&\leqslant\frac{1}{n}(\sum_{k\notin\{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}\}}|l(y_k,f_S^A(v_k))-l(y_k,\\
&\quad f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v_k))|+l(y_{i_1},f_S^A(v_{i_1}))-l(y'_{i_1},\\
&\quad f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{i_1}))|+|l(y_{i_2},f_S^A(v_{i_2}))-l(y'_{i_2},\\
&\quad f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{i_2}))|\cdots+|l(y_{i_{k_1}},f_S^A(v_{i_{k_1}}))-\\
&\quad l(y'_{i_{k_1}},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{i_{k_1}}))|+|l(y_{j_1},f_S^A(v_{j_1}))-\\
&\quad l(y'_{j_1},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{j_1}))|+|l(y_{j_2},f_S^A(v_{j_2}))-\\
&\quad l(y'_{j_2},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{j_2}))|\cdots+|l(y_{j_{k_2}},f_S^A(v_{j_{k_2}}))-\\
&\quad l(y'_{j_{k_2}},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{j_{k_2}}))|)\\
&\leqslant\frac{1}{n}(\sum_{k\notin\{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}\}}|l(y_k,f_S^A(v_k))-l(y_k,\\
&\quad f^A_{S^{\setminus\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}}(v_k))|+\sum_{k\notin\{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}\}}|\\
&\quad l(y_k,f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v_k))-l(y_k,f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}\\
&\quad(v_k))|+|l(y_{i_1},f_S^A(v_{i_1}))-l(y'_{i_1},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}
\end{aligned}$$

$$(v'_{i_1}))|+|l(y_{i_2},f_S^A(v_{i_2}))-l(y'_{i_2},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{i_2}))|\cdots+|l(y_{i_{k_1}},f_S^A(v_{i_{k_1}}))-l(y'_{i_{k_1}},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{i_{k_1}}))|+|l(y_{j_1},f_S^A(v_{j_1}))-l(y'_{j_1},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{j_1}))|+|l(y_{j_2},f_S^A(v_{j_2}))-l(y'_{j_2},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{j_2}))|\cdots+|l(y_{j_{k_2}},f_S^A(v_{j_{k_2}}))-l(y'_{j_{k_2}},f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(v'_{j_{k_2}}))|)$$

$$\leqslant 2\Xi_{k_1,k_2}(n)+\frac{(k_1+k_2)M}{n}$$

同上叙述,$S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}$ 和 $S^{\backslash\{i_1,i_2,\cdots,i_{k_1}\},\{j_1,j_2,\cdots,j_{k_2}\}}$ 这两个样本集合在 $j_1,j_2,\cdots,j_{k_2}$ 位置上替换了不同的本体样本,从而确保了最后一个不等式成立。

进而有:

$$|\Phi_A(S)-\Phi_A(S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}})|$$

$$=|(R(f_S^A)-\hat{R}_S(f_S^A))-(R(f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}})-\hat{R}_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}))|$$

$$\leqslant|R(f_S^A)-R(f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}})|+|\hat{R}_S(f_S^A)-\hat{R}_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}}(f^A_{S^{i_1,i_2,\cdots,i_{k_1},j_1,j_2,\cdots,j_{k_2}}})|$$

$$\leqslant 4\Xi_{k_1,k_2}(n)+\frac{(k_1+k_2)M}{n}$$

将其与前文的引理 1 相结合,即得引理 10。

引理 11 刻画了在 $\Xi_{k_1,k_2}(n)$组合稳定条件下 $\Phi_A(S)$的期望值的上界。

引理 11 设非负整数 k_1,k_2,n 满足 $0<k_1+k_2\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$,$G$ 是 S 的决定图。假设本体学习算法 A 是 $\Xi_{k_1,k_2}(i)$稳定的($i\leqslant n$ 是任意的),且图 G 的最大度为 Δ。设 $\Xi_{k_1,k_2}(n,\Delta)=\max_{i\in[0,\Delta]}\Xi_{k_1,k_2}(n-i)$,有:

$$E[\Phi_A(S)]\leqslant 2(\Delta+1)\Xi_{k_1,k_2}(n,\Delta)$$

证明:首先需要证明,在引理 11 组合稳定条件成立的情况下,有:

$$\max_{(v_j,y_j)\in S}E_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]\leqslant 2(\Delta+1)\Xi_{k_1,k_2}(n,\Delta)$$

对于任意 $i\in\{1,\cdots,n\}$,设 $N_G^+(i)=\{j_1,\cdots,j_{n_i}\}$满足 $j_{k-1}>j_k$。对任意 $k\in\{1,\cdots,n_i\}$定义 $S^{(i,-k_1-k_2+1)}=S^{(i,-1)}=S^{(i,0)}=S$,且 $S^{(i,k)}$通过在 $S^{(i,k-k_1-k_2)}$中删除第 $j_k,j_{k-1},\cdots,j_{k-k_1+1}$个(共 k_1 个)本体样本后,再替换第 $j_{k-k_1},j_{k-k_1-1},\cdots,j_{k-k_1-k_2+1}$个(共 k_2 个)本体样本得到。根据本体算法 A 的稳定性,假设有:

$$|l(y,f^A_{S^{(i,k-k_1-k_2)}}(v))-l(y,f^A_{S^{(i,k)}}(v))\leqslant\Xi_{k_1,k_2}(n,\Delta)$$

同时注意到:

$$l(y,f_S^A(v))=\sum_{k=1}^{n_i}(l(y,f^A_{S^{(i,k-k_1-k_2)}}(v))-l(y,f^A_{S^{(i,k)}}(v)))+l(y,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v))+\cdots+l(y,f_S^A(v_i))$$

$$l(y_i,f_S^A(v_i))=\sum_{k=1}^{n_i}(l(y_i,f^A_{S^{(i,k-k_1-k_2)}}(v))-l(y,f^A_{S^{(i,k)}}(v)))+l(y_i,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v_i))+\cdots+l(y_i,f^A_{S^{(i,n_i)}}(v_i))$$

因此,得到:

$$l(y,f_S^A(v))-l(y_i,f_S^A(v_i))$$

$$=\sum_{k=1}^{n_i}(l(y,f^A_{S^{(i,k-k_1-k_2)}}(v))-l(y,f^A_{S^{(i,k)}}(v)))-\sum_{k=1}^{n_i}(l(y_i,f^A_{S^{(i,k-k_1-k_2)}}(v))-l(y,f^A_{S^{(i,k)}}(v)))+(l(y,f^A_{S^{(i,n_i)}}(v))-l(y_i,f^A_{S^{(i,n_i)}}(v_i)))+(l(y,f^A_{S^{(i,n_i-1)}}(v))-l(y_i,f^A_{S^{(i,n_i-1)}}(v_i)))+\cdots+(l(y,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v))-l(y_i,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v_i)))$$

$$\leqslant\sum_{k=1}^{n_i}|l(y,f^A_{S^{(i,k-k_1-k_2)}}(v))-l(y,f^A_{S^{(i,k)}}(v))|+\sum_{k=1}^{n_i}|l(y_i,f^A_{S^{(i,k-k_1-k_2)}}(v))-l(y,f^A_{S^{(i,k)}}(v))|+(l(y,f^A_{S^{(i,n_i)}}(v))-l(y_i,f^A_{S^{(i,n_i)}}(v_i)))+(l(y,f^A_{S^{(i,n_i-1)}}(v))-l(y_i,f^A_{S^{(i,n_i-1)}}(v_i)))+\cdots+(l(y,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v))-l(y_i,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v_i)))$$

$$\leqslant 2n_i\Xi_{k_1,k_2}(n,\Delta)+(l(y,f^A_{S^{(i,n_i)}}(v))-l(y_i,f^A_{S^{(i,n_i)}}(v_i)))+(l(y,f^A_{S^{(i,n_i-1)}}(v))-l(y_i,f^A_{S^{(i,n_i-1)}}(v_i)))+\cdots+(l(y,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v))-l(y_i,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v_i)))$$

由于(v_i,y_i)和(v,y)关于 $S^{(i,n_i)}$,$S^{(i,n_i-1)}$,…,$S^{(i,n_i-k_1-k_2+1)}$都是独立的,可得:

$$E_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]$$

$$=E_{S,(v,y)}[l(y,f^A_{S^{(i,n_i)}}(v))-l(y_i,f^A_{S^{(i,n_i)}}(v_i))+l(y,f^A_{S^{(i,n_i-1)}}(v))-l(y_i,f^A_{S^{(i,n_i-1)}}(v_i))+\cdots+l(y,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v))-l(y_i,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v_i))]+2n_i\Xi_{k_1,k_2}(n,\Delta)$$

$$\leqslant E_{S,(v,y)}[l(y,f^A_{S^{(i,n_i)}}(v))-l(y_i,f^A_{S^{(i,n_i)}}(v_i))+l(y,f^A_{S^{(i,n_i-1)}}(v))-l(y_i,f^A_{S^{(i,n_i-1)}}(v_i))+\cdots+l(y,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v))-l(y_i,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v_i))]+2(\Delta+1)\Xi_{k_1,k_2}(n,\Delta)$$

$$=E_{S,(v,y)}[l(y,f^A_{S^{(i,n_i)}}(v))]-E_S[l(y_i,f^A_{S^{(i,n_i)}}(v_i))]+E_{S,(v,y)}[l(y,f^A_{S^{(i,n_i-1)}}(v))]-E_S[l(y_i,f^A_{S^{(i,n_i-1)}}(v_i))]+\cdots+E_{S,(v,y)}[l(y,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v))]-E_S[l(y_i,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v_i))]+2(\Delta+1)\Xi_{k_1,k_2}(n,\Delta)$$

$$=E_{S^{(i,n_i)},(v,y)}[l(y,f^A_{S^{(i,n_i)}}(v))]-E_{S^{(i,n_i)},(v_i,y_i)}[l(y_i,f^A_{S^{(i,n_i)}}(v_i))]+E_{S^{(i,n_i-1)},(v,y)}[l(y,f^A_{S^{(i,n_i-1)}}(v))]-E_{S^{(i,n_i-1)},(v_i,y_i)}[l(y_i,f^A_{S^{(i,n_i-1)}}(v_i))]+\cdots+E_{S^{(i,n_i-k_1-k_2+1)},(v,y)}[l(y,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v))]-E_{S^{(i,n_i-k_1-k_2+1)},(v_i,y_i)}[l(y_i,f^A_{S^{(i,n_i-k_1-k_2+1)}}(v_i))]+2(\Delta+1)\Xi_{k_1,k_2}(n,\Delta)$$

$$=2(\Delta+1)\Xi_{k_1,k_2}(n,\Delta)$$

最后,有:

$$E[\Phi_A(S)]=E_S[E_{(v,y)}[l(y,f_S^A(v))]-\frac{1}{n}\sum_{i=1}^{n}l(y_i,f_S^A(v_i))]$$

$$=\frac{1}{n}\sum_{i=1}^{n}E_{S,(v,y)}[l(y,f_S^A(v))-l(y_i,f_S^A(v_i))]$$

$$\leqslant 2(\Delta+1)\Xi_{k_1,k_2}(n,\Delta)$$

结合引理 10 和引理 11,可得到$\Xi_{k_1,k_2}(n)$组合稳定条件下本体学习算法的广义界如下。

定理 9 设非负整数 k_1,k_2,n 满足 $0<k_1+k_2\ll n$。给定 n 个本体顶点的本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$,$G$ 是 S 的决定图。假设本体学习算法 A 是 $\Xi_{k_1,k_2}(i)$稳定的($i\leqslant n$ 是任意的),本体亏损函数的上界为 M,且图 G 的最大度为 Δ。

设 $\Xi_{k_1,k_2}(n,\Delta)=\max_{i\in[0,\Delta]}\Xi_{k_1,k_2}(n-i)$。则对任意 $\delta\in(0,1)$,下式以置信度 $1-\delta$ 成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(\Delta+1)\Xi_{k_1,k_2}(n,\Delta)+\frac{4n\Xi_{k_1,k_2}(n)+(k_1+k_2)M}{n}\sqrt{\frac{\Lambda(G)\ln(\frac{1}{\delta})}{2}}$$

更进一步,若本体样本集满足 m-独立的特殊条件,本节的最后结果刻画了在 $\Xi_{k_1,k_2}(n)$ 组合稳定条件下,广义界可以做如下修改。

定理 10 设非负整数 k_1,k_2,n 满足 $0<k_1+k_2\ll n$。给定 n 个本体顶点的 m-独立本体样本 $S=\{(v_1,y_1),\cdots,(v_n,y_n)\}$,$G$ 是 S 的决定图。假设本体学习算法 A 是 $\Xi_{k_1,k_2}(i)$ 稳定的($i\leqslant n$ 是任意的),本体亏损函数的上界为 M,则对任意 $\delta\in(0,1)$,下式以置信度 $1-\delta$ 成立。

$$R(f_S^A)\leqslant\hat{R}(f_S^A)+2(m+1)\Xi_{k_1,k_2}(n,2m)+(4n\Xi_{k_1,k_2}(n)+(k_1+k_2)M)\sqrt{\frac{2m\ln(\frac{1}{\delta})}{n}}$$

结束语 从本体算法稳定性入手,首先对 PO 一致稳定条件和 LTO 一致稳定条件下的本体算法的广义界进行了分析;再针对大样本本体算法给出了 Pk 一致稳定性和 LkO 一致稳定性的定义,并由此给出了对应的广义界;最后给出组合条件一致稳定性定义,并推导出最终的大样本条件下以及广义样本变化稳定条件下,本体学习算法的广义界。

特别需要说明第 2 节与第 5 节和第 6 节定义的本质区别。虽然之后的定义是前面替换单个本体样本以及删除两个样本的稳定性的扩展,但是 PO 一致稳定性和 LTO 一致稳定性适合所有样本容量的本体算法,是一般意义下的稳定性定义。一旦推广到后者,必须是大样本本体算法的情景,Pk 一致稳定性、LkO 一致稳定性以及组合条件一致稳定性的定义才有意义,即要求样本替换和删除的本体样本数量相对于总样本容量而言是微不足道的。即,第 3 节、第 4 节中叙述的理论结果适用于所有本体算法,但第 5 节、第 6 节中的定义和理论结果只适用于大样本本体算法。

参 考 文 献

[1] PALOMBI O, JOUANOT F, NZIENGAM N, et al. OntoSIDES: Ontology-based student progress monitoring on the national evaluation system of French Medical Schools[J]. Artificial Intelligence in Medicine, 2019, 96: 59-67.

[2] ARBABI A, ADAMS D R, FIDLER S, et al. Identifying clinical terms in medical text using ontology-guided machine learning [J]. JMIR Medical Informatics, 2019, 7(2): 191-205.

[3] MESSAOUDI R, JAZIRI F, MTIBAA A, et al. Ontology-based approach for liver cancer diagnosis and treatment[J]. Journal of Digital Imaging, 2019, 32(1): 116-130.

[4] THOMAS P D, HILL D P, MI H Y, et al. Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems[J]. Nature Genetics, 2019, 51(10): 1429-1433.

[5] ZHAO Y W, FU G Y, WANG J, et al. Gene function prediction based on Gene Ontology Hierarchy Preserving Hashing[J]. Genomics, 2019, 111(3): 334-342.

[6] HADAROVICH A, ANISHCHENKO I, TUZIKOV A V, et al. Gene ontology improves template selection in comparative protein docking[J]. Proteins-Structure Function and Bioinformatics, 2019, 87(3): 245-253.

[7] TANG H, FINN R D, THOMAS P D. TreeGrafter: phylogenetic tree-based annotation of proteins with Gene Ontology terms and other annotations[J]. Bioinformatics, 2019, 35(3): 518-520.

[8] MAO Z T, MA H W. iMTBGO: An algorithm for integrating metabolic networks with transcriptomes based on gene ontology analysis[J]. Current Genomics, 2019, 20(4): 252-259.

[9] VISSER U, STUCKENSCHMIDT H, SCHUSTER G, et al. Ontologies for geographic information processing[J]. Computers & Geosciences, 2002, 28(1): 103-117.

[10] FONSECA F, EGENHOFER M, DAVIS C, et al. Semantic granularity in ontology-driven geographic information systems [J]. Annals of Mathematics and Artificial Intelligence, 2002, 36(1/2): 121-151.

[11] ZHU L L, HUA G, ASLAM A. Ontology learning algorithm using weak functions[J]. Open Physics, 2018, 16: 910-916.

[12] ZHU L L, HUA G, ZAFAR S, et al. Fundamental ideas and mathematical basis of ontology learning algorithm[J]. Journal of Intelligent and Fuzzy Systems, 2018, 35(4): 4503-4516.

[13] XIAO Z H, YU H, WANG Y C. Unsupervised Machine Learning Based on the Sum of Squares of the Dynamic Deviations[J]. Journal of Chongqing University of Technology (Natural Science), 2018, 32(11): 134-139.

[14] YE S R, SUN N. Chinese Text Classification by Domain Ontology Graph Based on Concept Clustering [J]. Computer Engineering, 2016, 42(12): 181-187.

[15] GAO W, XU T W. Stability analysis of learning algorithms for ontology similarity computation[J]. Abstract and Applied Analysis, 2013.

[16] GAO W, ZHANG Y G, XU T W, et al. Analysis for learning a similarity function with ontology applications[J]. Journal of Information & Computational Science, 2012, 17(9): 5311-5318.

[17] CHEN L H Y. Two central limit problems for dependent random variables[J]. Probability Theory and Related Fields, 1978, 43(3): 223-243.

[18] BALDI P, RINOTT Y. On normal approximations of distributions in terms of dependency graphs[J]. The Annals of Probability, 1989, 17(4): 1646-1650.

[19] ZHANG R, LIU X W, WANG Y Y, et al. McDiarmid-Type Inequalities for Graph-Dependent Variables and Stability Bounds [C]// Thirty-third Conference on Neural Information Processing Systems. Vancouver, Canada. Science, 2018: 1-20.

ZHU Lin-li, born in 1975, Ph.D, senior engineer. His main research interests include artificial intelligence and machine learning.