ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.20, No.11, November 2009, pp.2965−2976
doi: 10.3724/SP.J.1001.2009.03473

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

从不确定图中挖掘频繁子图模式*

邹兆年, 李建中+, 高 宏, 张 硕

(, 150001)

Mining Frequent Subgraph Patterns from Uncertain Graphs

ZOU Zhao-Nian, LI Jian-Zhong+, GAO Hong, ZHANG Shuo

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: E-mail: lijzh@hit.edu.cn, http://db.hit.edu.cn

Zou ZN, Li JZ, Gao H, Zhang S. Mining frequent subgraph patterns from uncertain graphs. *Journal of Software*, 2009,20(11):2965−2976. http://www.jos.org.cn/1000-9825/3473.htm

Abstract: This paper studies uncertain graph data mining and especially investigates the problem of mining frequent subgraph patterns from uncertain graph data. A data model is introduced for representing uncertainties in graphs, and an expected support is employed to evaluate the significance of subgraph patterns. By using the apriori property of expected support, a depth-first search-based mining algorithm is proposed with an efficient method for computing expected supports and a technique for pruning search space, which reduces the number of subgraph isomorphism testings needed by computing expected support from the exponential scale to the linear scale. Experimental results show that the proposed algorithm is 3 to 5 orders of magnitude faster than a naïve depth-first search algorithm, and is efficient and scalable.
Key words: uncertain graph; graph mining; frequent subgraph pattern

摘 要: YN V'... ut
FY[g Apriori S{t n
,8
Y[gN .r 8 8¥
Ù 3~5 Q ‰
关键词:
中图法分类号: TP311 文献标识码: A

, ( ),
. , . [1−7].
( ). , ,

---

* Supported by the National Natural Science Foundation of China under Grant Nos.60533110, 60773063 ( ) the National Basic Research Program of China under Grant No.2006CB303005 ((973)); the Program for New Century Excellent Talents in University of China under Grant No.NCET-05-0333 ( )

Received 2008-05-29; Accepted 2008-10-09; Published online 2008-12-22

, (protein-protein interaction, PPI) , ,
. PPI , PPI . [8] PPI
PPI . 1 [8] PPI , ,
, PPI . , PPI .

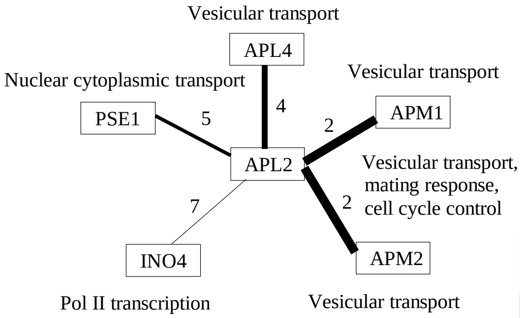


Fig.1 An example of PPI network

1

. , .
PPI , . ,
. , , .
, ,
. , .
, ,
. . ,
. I G , I G I G .
. $d=\{I_1,I_2,\ldots,I_n\}$
$D=\{G_1,G_2,\ldots,G_n\}$, $1\le i\le n$ G_i I_i. 2 .
. ,
. , ,
. , S D S D
. ,
. . *minsup*,
. , : D *minsup*, D
.
:
(1) . . ,
, .
(2) D D
. D , . ,
.
, . :
(1) Apriori , .
Apriori .

(2) .

, .

(3) .

.

(4) .

1 . 2 .

3 . 4 . 5 .

1 相关工作

. [1,2] AGM

FSG . [3−5] gSpan FFSM Gaston .

, [6] CloseGraph . [7] SPIN

. , .

[9] [10] Top-k [11] Skyline

[12]. , . , [13]

Apriori U-Apriori . [14] [13] . [15]

. , .

2 问题定义

定义 1. $G=((V,E),\Sigma,L,P)$, (V,E) ,V ,E ,Σ ,

$L:V\cup E\rightarrow\Sigma$, $P:E\rightarrow(0,1]$.

.1 . ,

1 , $((V,E),\Sigma,L)$.

. $I=((V',E'),\Sigma',L')$ $G=((V,E),\Sigma,L,P)$ ($G\Rightarrow I$), $V'=V,E'\subseteq E,\Sigma'=\Sigma,L'=L|_{V'\cup E'}$,

$L|_{V'\cup E'}$ L $V'\cup E'$. ,

. , $G=((V,E),\Sigma,L,P)$ $I=((V',E'),\Sigma',L')$

$$P(G\Rightarrow I)=\prod\nolimits_{e\in E'}P(e)\cdot\prod\nolimits_{e\in E-E'}(1-P(e)) \tag{1}$$

(1) E' I , $E-E'$ I . $Imp(G)$ G

. G ,$Imp(G)$ $2^{|E|}$ . .

定理 1. G, $P(G\Rightarrow I)$ $Imp(G)$.

. .

$d=\{I_i|1\le i\le n\}$ $D=\{G_i|1\le i\le n\}$ ($D\Rightarrow d$), $1\le i\le n$ $G_i\Rightarrow I_i$. $Imp(D)$

D . ,$Imp(D)$ $\prod_{i=1}^{n}2^{|E_i|}$.

, D d

$$P(D\Rightarrow d)=\prod\nolimits_{i=1}^{n}P(G_i\Rightarrow I_i) \tag{2}$$

定理 2. D, $P(D\Rightarrow d)$ $Imp(D)$.

例 1: 2(a) $D=\{G_1,G_2\}$, , G_1 v_1

A; , G_1 (v_1,v_2) x; , G_1

(v_1,v_2) 0.5. ,G_1 $2^4=16$ (3).G_2

$2^3=8$ .D 16×8=128 .

定义 2. $G=((V,E),\Sigma,L)$ $G'=((V',E'),\Sigma',L')$($G\subseteq_C G'$),

$f:V\to V'$:(1) $\forall v\in V, L(v)=L'(f(v))$;(2) $\forall (u,v)\in E,(f(u),f(v))\in E'$;(3) $\forall (u,v)\in E, L((u,v))=L'((f(u),f(v)))$.$G'$ (V'',E'') f G G' , ,$V''=\{f(v)|v\in V\}, E''=\{(f(u),f(v))|(u,v)\in E\}$.

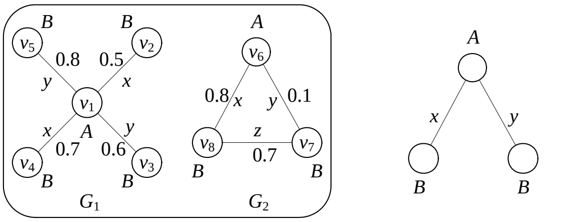


Fig.2 An example of uncertain graph database

2

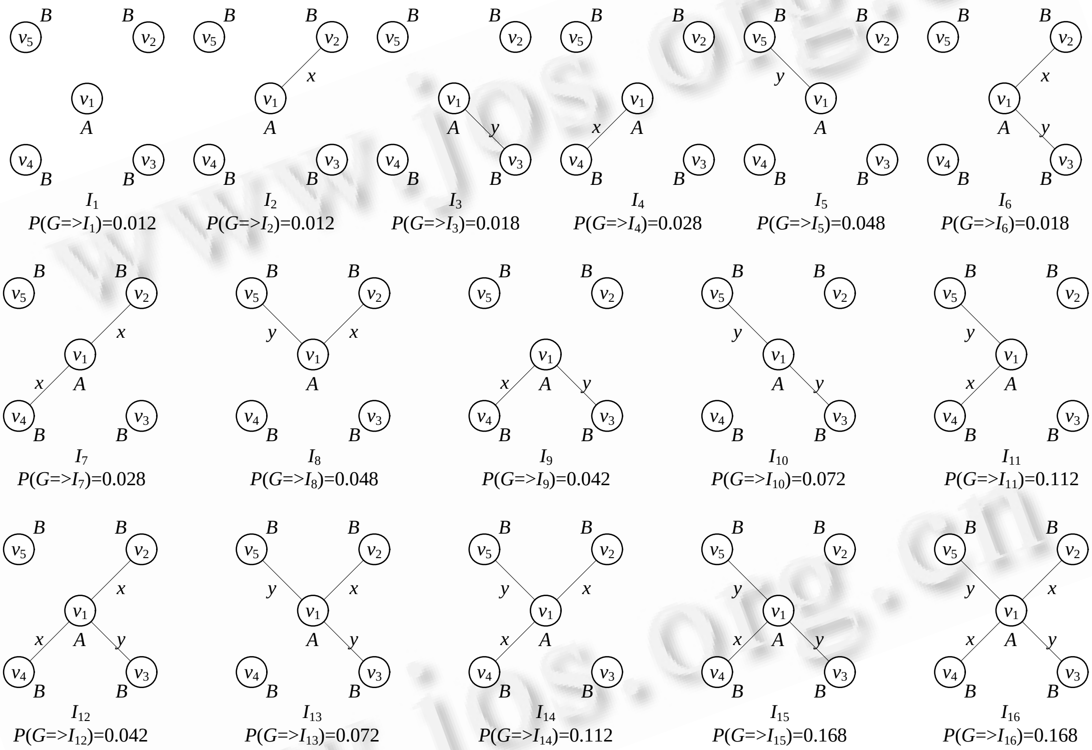


Fig.3 Probability distribution of all certain graphs implicated by uncertain graph G_1 in Fig.2

3 2 G_1

, , 1 . S D $sup_D(S)=|\{G|S\subseteq_C G, G\in D\}|/|D|$. , , . , .

定义 3. S D , S D 1 G . S D k, S D k- . S S', $S\subseteq_C S'$, S S' ,S' S .S S' S S' $|E_S|+1=|E_{S'}|$.

定义 4. D, $Imp(D)$ D , S D

:

$$\begin{bmatrix} s_1 & s_2 & \dots & s_m \\ P(s_1) & P(s_2) & \dots & P(S_m) \end{bmatrix},$$

$,s_1,s_2,\dots,s_m$ S $Imp(D)$ $,P(s_i)=\sum_{d\in Imp(D)|sup_d(S)=s_i} P(D\Rightarrow d)$ s_i

$,m=|\{sup_d(S)|d\in Imp(D)\}|$.

定义 5. S D 4 , S D

$$esup_D(S)=\sum_{i=1}^{m} s_i\cdot P(s_i)=\sum_{d\in Imp(D)} sup_d(S)\cdot P(D\Rightarrow d) \quad (3)$$

S D , S D $minsup\in[0,1]$. ,

: D $minsup$, D

$FP=\{S|S$ D , $esup_D(S)\geq minsup\}$.

S G ($S\subseteq_U G$), S 1 G . ,S G

$$P(S\subseteq_U G)=\sum_{I\in Imp(G)} P(G\Rightarrow I)\cdot\psi(I,S) \quad (4)$$

$,\psi(I,S)=1$, $S\subseteq_C I$; $,\psi(I,S)=0$. , (4) (3)

$$esup_D(S)=\frac{1}{|D|}\cdot\sum_{i=1}^{|D|} P(S\subseteq_U G_i) \quad (5)$$

(5) S D , (5) $|D|$ (3)

$|Imp(D)|$. (4) (5), :

(1) G, G Apriori , S S', S S'

, $P(S\subseteq_U G)\geq P(S'\subseteq_U G)$;

(2) D, D Apriori , S S', S

S' , $esup_D(S)\geq esup_D(S')$.

Apriori , ,

. .

3 频繁子图模式挖掘算法

3.1 算法概述

D $minsup.D$ (3)

, D , , D

$.D$ k- k . 4 2 D

, , . 1 1-

$e_1,e_2,\dots,e_k$, k , , i

e_i, $e_1,e_2,\dots,e_{i-1}$. 4 .

D ,

.

. :

, 1 1- .

S, S D $esup_D(S)$, $esup_D(S)\geq minsup$, S , S

S (S); $esup_D(S)<minsup$, S , Apriori ,S

, S

S . 1 1-

, . , .

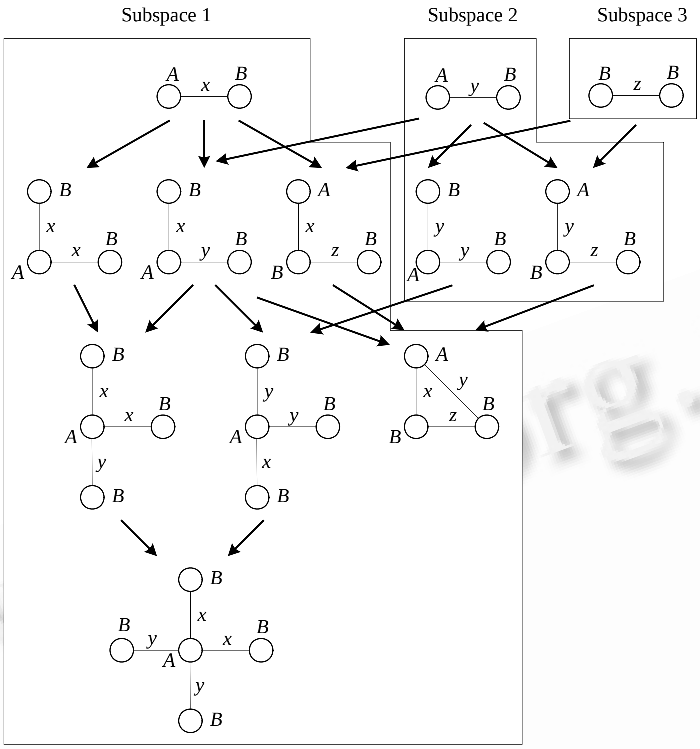


Fig.4 Search space of subgraph patterns

4

, , S ,

S . S , S

, S , S .

, . 3.2

. 3.3 . 3.4 .

3.2 计算期望支持度的高效算法

(5), S D $esup_D(S)$ S D

G_i $P(S\subseteq_U G_i)$. (4), $P(S\subseteq_U G_i)$ G_i $2^{|E_{G_i}|}$

, S G_i . ,

$\sum_{i=1}^{|D|}2^{|E_{G_i}|}$. , .

S D $|D|$.

3.2.1

$G\Rightarrow\Gamma(S_1,S_2,\ldots,S_k)$: G $\Gamma(S_1,S_2,\ldots,S_k)$, S_1,

$S_2,\ldots,S_k$. (1) $G\Rightarrow\Gamma(S_1,S_2,\ldots,S_k)$

$$P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_k))=\sum_{I\in\Gamma(S_1,S_2,\ldots,S_k)}P(G\Rightarrow I)=\prod_{e\in E_{S_1}\cup E_{S_2}\cup\ldots\cup E_{S_k}}P(e) \quad (6)$$

.

定理 3. G S. S G $\{S_1,S_2,\ldots,S_l\}$, S G

$$P(S\subseteq_U G)=\sum_{1\le i\le \ell}P(G\Rightarrow \Gamma(S_i))-\sum_{1\le i_1<i_2\le \ell}P(G\Rightarrow \Gamma(S_{i_1},S_{i_2}))+\ldots+(-1)^{j-1}\sum_{1\le i_1<i_2<\ldots<i_j\le \ell}P(G\Rightarrow \Gamma(S_{i_1},S_{i_2},\ldots,S_{i_j}))+\ldots+(-1)^{\ell-1}\sum_{1\le i_1<i_2<\ldots<i_\ell\le \ell}P(G\Rightarrow \Gamma(S_{i_1},S_{i_2},\ldots,S_{i_\ell}))\qquad(7)$$

定理 4. $S_1,S_2,\ldots,S_k$ S G k . S_p S_q $(1\le p<q\le k)$, $E_{S_p}\cap E_{S_q}=\varnothing$,

$$P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_k))=P_1\cdot P_2/P_3\qquad(8)$$

, $P_1=P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_{p-1},S_{p+1},\ldots,S_k))$,

$P_2=P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_{q-1},S_{q+1},\ldots,S_k))$,

$P_3=P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_{p-1},S_{p+1},\ldots,S_{q-1},S_{q+1},\ldots,S_k))$.

3 4, S G $P(S\subseteq_U G)$. , , . , S G EM, $EG=(V_{EG},E_{EG})$, V_{EG} EM , E_{EG} EM , $S_i,S_j\in V_{EG}$ $(S_i,S_j)\in E_{EG}$ $E_{S_i}\cap E_{S_j}\neq\varnothing$. ,$EM$ S_i S_j $(S_i,S_j)\notin E_{EG}$. :

算法 1. COMP-OCC-PROB.

: S G.

:S G $P(S\subseteq_U G)$.

步骤 1. S G EM EG.

步骤 2. k 1 $|EM|$, EM k $\{S_1,S_2,\ldots,S_k\}$ $P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_k))$.

(1) $k=1$, 1 $\{S_i\}\subseteq EM$, (6) $P(G\Rightarrow\Gamma(S_i))$.

(2) $k=2$, 2 $\{S_i,S_j\}\subseteq EM$, S_i S_j EG , S_i S_j , $P(G\Rightarrow\Gamma(S_i,S_j))=P(G\Rightarrow\Gamma(S_i))\cdot P(G\Rightarrow\Gamma(S_j))$, , (6) $P(G\Rightarrow\Gamma(S_i,S_j))$.

(3) $3\le k\le|EM|$, k $\{S_1,S_2,\ldots,S_k\}\subseteq EM$, $S_p,S_q\in\{S_1,S_2,\ldots,S_k\}$, (8) $P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_k))$, , (6) $P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_k))$.

步骤 3. (7) S G $P(S\subseteq_U G)$ $P(S\subseteq_U G)$.

1, . .

算法 2. COMP-EXP-SUP.

: S D.

:S D $esup_D(S)$.

步骤 1. D G_i, 1 S G_i $P(S\subseteq_U G_i)$.

步骤 2. (5) S D $esup_D(S)$ $esup_D(S)$.

3.2.2

3 4 (5) 1 2 . 2 , D G_i, 2 1 S G_i , S G_i . , 2 $|D|$. $\sum_{i=1}^{|D|}2^{|E_{Gi}|}$, .

引理 1. S G $\{S_1,S_2,\ldots,S_k\}$. $S_p,S_q\in\{S_1,S_2,\ldots,S_k\}$, 1 $O(1)$ $P(G\Rightarrow\Gamma(S_1,S_2,\ldots,S_k))$.

定理 5. G S. EG S G . EG n

m , (7) $O(1)$ $P(G \Rightarrow \Gamma(S_{i_1}, S_{i_2}, ..., S_{i_j}))$ $1-(2^n-1)^{-1}(2^{n'}+2^{m'}+n-n'-2)$. ,$n'$ m' , $\binom{n'}{2} \le m, \binom{n'+1}{2} > m, \binom{n'}{2} + m' = m$.

5 EG n=10 , (7) $O(1)$ EG m . 5 , . , EG m=20 , $1-(2^{10}-1)^{-1}\times(2^6+2^5+10-5-2)\approx90\%$. , 1 .

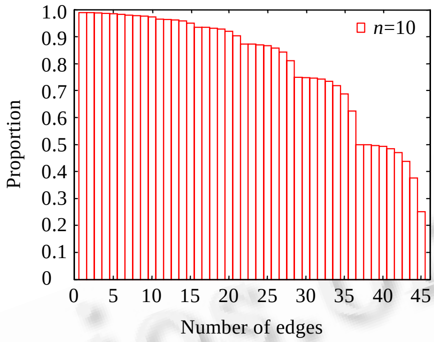


Fig.5 Relationship between the proportion of terms in Eq.(7) that can be computed in $O(1)$ time and the number of edges in EG

5 (7) $O(1)$ EG

, , , EG

, m $\binom{n}{2}$, m n . $\binom{n'}{2} \le m$, $n' = O(\sqrt{n})$, $n<m$, $2^{n'} = 2^{O(\sqrt{n})} > n$. , $O(1-2^{-\sqrt{n}})$, (7) $O(1)$.

3.3 子图模式搜索空间裁剪技术

, S $esup_D(S)$ $minsup$, Apriori ,S , . 2 , S , S G_i $P(S \subseteq_U G_i)$. , . $P(S \subseteq_U G_i)$ S. :

定理 6. D. S S' S' S , k=0,1,…,|D|,

$$esup_D(S) \le \frac{1}{|D|}\left(\sum_{i=1}^{k} P(S \subseteq_U G_i) + \sum_{i=k+1}^{|D|} P(S' \subseteq_U G_i)\right) \quad (9)$$

(9) S . (9) $minsup$, S . , (9) $esup_D(S)$ S . , (9) $P(S' \subseteq_U G_i)$ S , i=1,2,…,k $P(S \subseteq_U G_i)$. 2, . :

算法 3. COMP-UPP-SUP.

: S D $minsup$.

:S D $esup_D(S)$.

步骤 1. k=1,$upper$=0.S'= S .

步骤 2. k>|D|, $upper$, ; , 1 S G_k $P(S \subseteq_U G_k)$, (9) , $esup_D(S)$ $upper$.

步骤 3. $upper$<$minsup$, $upper$, ; ,k=k+1 2 .

, *S* , 3 2 . (5), $upper=esup_D(S)$.

3.4 完整算法

3.2 3.3 3.1 , .

算法 4. MINE-FREQ-SUBG.

: *D* *minsup*.

:*D* *FP*.

1. *FP* ;
2. *D* *D* 1- ;
3. **FOR** *D* 1- *S* **DO**
4. *DFS*(*S*,*D*);
5. *FP*;

4 . *DFS*(*S*,*D*).

6. 3 *S* *D* $esup_D(S)$ *upper*;
7. **IF** *upper*<*minsup* **THEN** ;
8. $FP=FP\cup\{S\}$;
9. *D* *S* ;
10. **FOR** *S* *S′* **DO**
11. **IF** *S′* **THEN** *DFS*(*S′*,*D*).

4 : , *FP* . *D* *D* 1- . 1- *S*, *DFS* *S* , . , *FP*.

DFS . : 3 *S* *D* $esup_D(S)$ *upper*. *upper*<*minsup*, . *S* , *S* . , *S* *FP* , *D* *S* . *S* *S′*, *S′* , *DFS*(*S′*,*D*), *S′* , . , . , . (DFS [3] CAM [4]) .

4 实 验

.

C . Intel Core2 Duo 2GHz CPU 2GB , Windows XP .

:

步骤 1. [2] . 6 :*D*( ),*V*( ),*E*( ),*I*( ),*L*( ), *T*( ).

步骤 2. 1 . *m* d^2 .

1 , Apriori NAÏVE, 6 4 *minsup* NAÏVE .

, NAÏVE . 1. 1 “–” NAÏVE . , D_1 *D*=20,*L*=10,*V*=5,*E*=1,*I*=5,*T*=10,*m*=0.9,*d*=0.1; D_2 *D*=40,*L*=10,*V*=5,*E*=1,*I*=5,*T*=10,*m*=0.9,*d*=0.1; D_3 *D*=20,*L*=10,*V*=5,*E*=1,*I*=5,*T*=15,

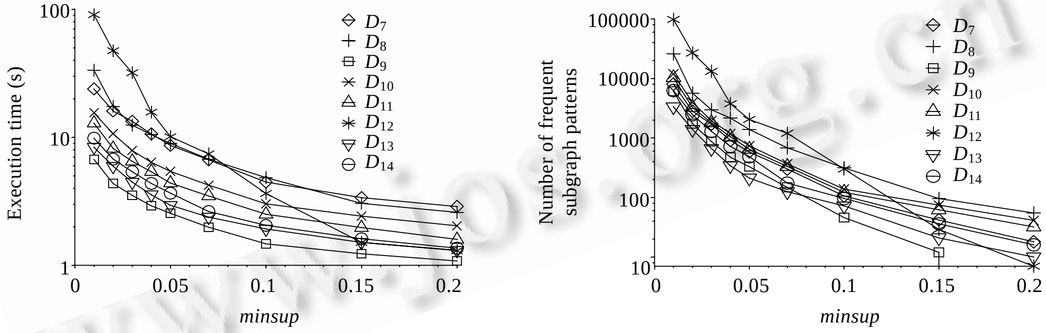
m=0.9,d=0.1; D_4 D=20,L=10,V=5,E=1,I=5,T=10,m=0.8,d=0.1; D_5 D=20,L=10, V=1,E=10,I=5,T=10,m=0.9,d=0.1; D_6 D=20,L=10,V=5,E=1,I=7,T=10,m=0.9,d=0.1. , NAÏVE 3~5 . , .

Table 1 Execution time comparison between the proposed algorithm and the NAÏVE algorithm

表 1 NAÏVE

Datasets	*minsup* (%)	Execution time (s)		Datasets	*minsup* (%)	Execution time (s)	
		Our algorithm	Algorithm NAÏVE			Our algorithm	Algorithm NAÏVE
D_1	80	0.00	0.86	D_4	80	0.00	0.93
	40	0.00	5.80		40	0.00	4.00
	20	0.01	55.71		20	0.01	32.13
	10	0.03	171.61		10	0.02	123.05
D_2	80	0.00	3.20	D_5	80	0.00	0.22
	40	0.01	9.42		40	0.08	9.29
	20	0.01	65.57		20	0.25	36.45
	10	0.08	232.25		10	3.26	114.27
D_3	80	0.01	15.93	D_6	80	0.00	443.64
	40	0.01	53.48		40	0.03	3 865.90
	20	0.42	311.05		20	0.15	–
	10	0.53	2 635.54		10	0.45	–

2 , *minsup* . 8 D_7,…,D_{14} .D_7 D=20000,L=100,V=10,E=10,I=5,T=20,m=0.95,d=0.05.D_8 D=10000,L=100,V=10, E=10,I=5,T=30,m=0.95,d=0.05.D_9 D=10000,L=200,V=10,E=10,I=5,T=20,m=0.95,d=0.05.D_{10} D=10000,L=100,V=5,E=10,I=5,T=20,m=0.95,d=0.05.D_{11} D=10000,L=100,V=10,E=5,I=5,T=20,m=0.95, d=0.05.D_{12} D=10000,L=100,V=10,E=10,I=10,T=20,m=0.95,d=0.05.D_{13} D=10000,L=100,V=10, E=10,I=5,T=20,m=0.8,d=0.05.D_{14} D=10000,L=100,V=10,E=10,I=5,T=20,m=0.95,d=0.01. 6(a) . , *minsup* . *minsup* , (6(b)), , .

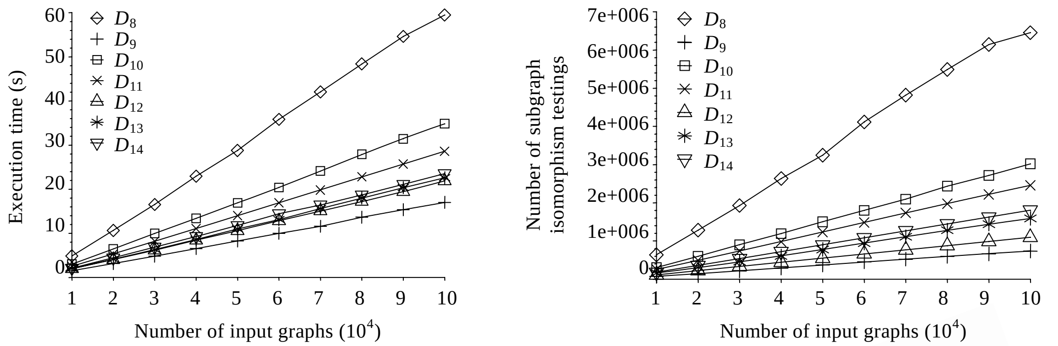


(a) (b)

Fig.6 Impact of the variation of *minsup* on the execution time of the algorithm

6 *minsup*

3 , . . 2 7 D_8,…,D_{14} . , *minsup* 10%. 7 . , . , .



(a) (b)

Fig.7 Impact of the variation of the number of input uncertain graphs on the scalability of the algorithm

7

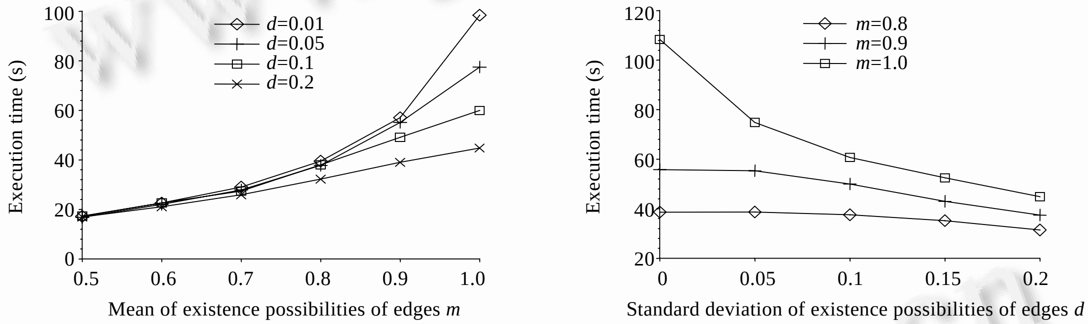
4 , .

4 . D=10000,L=100,V=10,E=10,I=5,T=40,d 0.01,0.05,0.1 0.2.

, *minsup* 1%. 8(a) . , ,

. , ,

, , .



(a) (b)

Fig.8 Impact of the variation of existence possibilities of edges on the execution time of the algorithm

8

4 , .

3 . D=10000,L=100,V=10,E=10,I=5,T=40,m 0.8,0.9 1.0. ,

minsup 1%. 8(b) . , ,

. , ,

, .

5 结 论

, .

Apriori , .

.

.

. , , .

References:

[1] Inokuchi A, Washio T, Motoda H. An apriori-based algorithm for mining frequent substructures from graph data. In: Zighed DA, Komorowski HJ, Zytkow JM, eds. Proc. of the 4th European Conf. on Principles of Data Mining and Knowledge Discovery. Lyon: Springer-Verlag, 2000. 13−23.

[2] Kuramochi M, Karypis G. Frequent subgraph discovery. In: Cercone N, Lin TY, Wu X, eds. Proc. of the 2001 IEEE Int'l Conf. on Data Mining. San Jose: IEEE Computer Society, 2001. 313−320.

[3] Yan X, Han J. gSpan: Graph-Based substructure pattern mining. In: Kumar V, Tsumoto S, Zhong N, Yu PS, Wu X, eds. Proc. of the 2002 IEEE Int'l Conf. on Data Mining. Maebashi: IEEE Computer Society, 2002. 721−724.

[4] Huan J, Wang W, Prins J. Efficient mining of frequent subgraphs in the presence of isomorphism. In: Kumar V, Tsumoto S, Zhong N, Yu PS, Wu X, eds. Proc. of the 2003 IEEE Int'l Conf. on Data Mining. Melbourne: IEEE Computer Society, 2002. 549−552.

[5] Nijssen S, Kok JN. A quickstart in frequent structure mining can make a difference. In: Kim W, Kohavi R, Gehrke J, DuMouchel W, eds. Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Seattle: ACM, 2004. 647−652.

[6] Yan X, Han J. Closegraph: Mining closed frequent graph patterns. In: Getoor L, Senator TE, Domingos P, Faloutsos C, eds. Proc. of the 9th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Washington: ACM, 2003. 286−295.

[7] Huan J, Wang W, Prins J, Yang J. Spin: Mining maximal frequent subgraphs from graph databases. In: Kim W, Kohavi R, Gehrke J, DuMouchel W, eds. Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Seattle: ACM, 2004. 581−586.

[8] Saito R, Suzuki H, Hayashizaki Y. Interaction generality, a measurement to assess the reliability of a protein-protein interaction. Nucleic Acids Research, 2002,30(5):1163−1168.

[9] Benjelloun O, Sarma AD, Halevy A, Theobald M, Widom J. Databases with uncertainty and lineage. VLDB Journal, 2008,17(2): 243−264.

[10] Dalvi N, Suciu D. Efficient query evaluation on probabilistic databases. VLDB Journal, 2007,16(4):523−544.

[11] Soliman MA, Ilyas IF, Chang KCC. Top-*k* query processing in uncertain databases. In: Chirkova R, Dogac A, Ozsu T, Sellis T, eds. Proc. of the 2002 IEEE Int'l Conf. on Data Engineering. Istanbul: IEEE Computer Society, 2007. 896−905.

[12] Pei J, Jiang B, Lin X, Yuan Y. Probabilistic skylines on uncertain data. In: Koch C, Gehrke J, Garofalakis MN, Srivastava D, Aberer K, Deshpande A, Florescu D, Chan CY, Ganti V, Kanne CC, Klas W, Neuhold EJ, eds. Proc. of the 24th Int'l Conf. on Very Large Data Bases. Austria: ACM, 2007. 15−26.

[13] Chui CK, Kao B, Hung E. Mining frequent itemsets from uncertain data. In: Zhou Z, Li H, Yang Q, eds. Proc. of the 11th Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining. Nanjing: Springer-Verlag, 2007. 47−58.

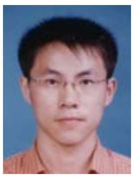
[14] Chui CK, Kao B. A decremental approach for mining frequent itemsets from uncertain data. In: Washio T, Suzuki E, Ting KM, Inokuchi I, eds. Proc. of the 12th Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining. Osaka: Springer-Verlag, 2008. 64−75.

[15] Zhang Q, Li F, Yi K. Finding frequent items in probabilistic data. In: Wang JT, ed. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. Vancouver: ACM, 2008. 819−832.

邹兆年(1979), , , , .

高宏(1966), , , , ,CCF , , .

李建中(1950), , , , ,CCF , , .



张硕(1982), , , .