人工知能学会 インタラクティブ
情報アクセスと可視化マイニング研究会(第24回)
SIG-AM-24-05

グラフ構造データを対象とした解釈可能決定集合の拡張

Interpretable Decision Sets for Graph Structured Databases

松山 航太 [1]　尾崎 知伸 [1] *
Kota Matsuyama[1]　Tomonobu Ozaki[1]

[1] 日本大学 文理学部
[1] College of Humanities and Sciences, Nihon University

Abstract: Interpretable Decision Sets (IDS) is one of representative algorithms for deriving classification rules having high interpretability while keeping high accuracy. In this paper, we propose an extension of IDS for handling graph structured data. In the proposal, after converting a graph database into a set of transactions using frequent subgraph mining, classification rules will be extracted based on our developed objective functions on interpretability of subgraph patterns.

1 はじめに

深層学習を中心とした機械学習技術の発展に伴い，近年，非常に高精度な機械学習モデルが構築されている．しかし機械学習モデルの多くは，利用者がその内容を理解・解釈することが困難な大規模かつ複雑なブラックボックスであり，モデルのデバッグ・更新が容易ではなく，また安全面や倫理面からもその応用範囲が限定されてしまうことも考えられる．これらの問題を軽減するため，モデル理解の支援や解釈可能（容易）なモデルの構築など，機械学習の解釈性が着目され，多数の研究が行われている [1, 2]．本研究では，分類規則抽出に関する代表的な解釈性研究の一つである解釈可能決定集合（Interpretable Decision Sets; IDS）[3] に着目する．IDS は，データから導出されるクラス相関ルール集合から，解釈性を考慮した評価関数を最大化する部分集合を抽出することで，分類精度を維持したまま解釈性の高い少数ルール群を特定する．本研究では，IDS のアイディアをグラフ構造データへと拡張する．具体的には，頻出部分グラフ発見技術 [4] を用いてグラフデータベースをトランザクション化した上で，クラス相関ルール [5] を導出する．加えて，クラス相関ルールの構成要素が部分グラフパターンであることに着目し，グラフ形状を考慮した解釈性に関する新たな評価尺度をルール集合選択基準に組み込む．これらにより，グラフ構造データからの解釈容易な分類規則集合の導出を目指す．

本論文の構成は以下の通りである．2 章で関連研究について言及する．3 章で IDS の要点を整理する．4 章で，グラフ構造データからのクラス相関ルールの導出およびグラフ構造を考慮したルール集合選択基準を導入し，IDS をグラフ構造データへと拡張する．5 章で評価実験を行い，最後に 6 章でまとめと今後の課題を述べる．

*連絡先：日本大学文理学部情報科学科
〒 156-8550 東京都世田谷区桜上水 3-25-40
E-mail: tozaki@chs.nihon-u.ac.jp

2 関連研究

近年，多様な側面から機械学習の解釈性に関する研究が活発に行われている [1, 2]．これらの研究は，入力事例に対する判断根拠を示すことを目的とした局所的説明に関する研究や，モデル全体を別の解釈容易なモデルで近似する大域的説明に関する研究などに大別することができる．本研究で対象とする IDS は，データから直接可読性の高い解釈可能モデルを構築する手法と認識されるが，実際にはクラス相関ルール集合から決定集合を抽出しており，モデルを経由する大域的説明に関する手法と関係が深い．

IDS と関連し，これまで，決定木アンサンブルとして表現されるモデルから少数のルールを抽出する手法が提案されている [6, 7]．なお，決定木中の根から各葉までのパスは一つのルールに対応するので，決定木アンサンブルからのルール抽出は，大規模ルール集合からルール抽出に相当する．例えば inTrees[6] は，決定木アンサンブルをルール集合に変換した後，不要な条件の削除と複数ルールの集約を行い，集合被覆アルゴリズムを用いて少数ルールを選択する．また defragTrees[7] は，分類器アンサンブルを確率的生成モデルとしてモデル化し，EM アルゴリズムやベイズ推定を用いた最適化を通じて少数ルールを導出する．

解釈性研究のグラフ構造データへの応用に関しては，文献 [8] において，グラフ構造データをトランザクショ

ン化した上で決定木アンサンブルを導出し，inTrees と defragTrees を用いて解釈容易なルールを抽出する試みが報告されている．一方，IDS の一つの拡張として，2 段階構造を持つ決定集合を対象に，元モデルへの忠実度と明確性，解釈容易性を同時に満たす少数ルールを選択する手法が開発されている [9]．

3　解釈可能決定集合

解釈可能決定集合は，下記の 2 段階処理を通じ，分類規則集合を獲得する．

1. クラス相関ルール集合の獲得：　アイテムの全体集合を $\mathcal{X}$，クラスラベルの全体集合を $\mathcal{C}$ とし，アイテム集合 $x_i \subseteq \mathcal{X}$ とそのクラスラベル $y_i \in \mathcal{C}$ の対 $t_i = \langle x_i, y_i \rangle$ をトランザクションとするデータベース（訓練例集合）$D = \{t_1, t_2, \cdots, t_{|D|}\}$ から，クラス相関ルール $r = (s, c) \in S \subseteq 2^{\mathcal{X}} \times C$ の集合 S を導出する．

2. 最適部分集合の獲得：　D と S を入力とし，（後述する）各評価関数 $f_i(D, R)$ $(i \in \{1, \cdots, 7\})$ の重み λ_i (> 0) を用いた線形和関数

$$F(D, R) = \sum_{i=1}^{7} \lambda_i f_i(D, R)$$

を最大化する S の部分集合

$$R^* = \arg\max_{R \subseteq S} F(D, R)$$

を最終的な決定集合（ルール集合）として出力する．ただし実際には，$F(D, R)$ が非単調な劣モジュラ関数であることに着目し，平滑化局所探索（Smooth Local Search; SLS）[10] を用いて R^* の近似解を獲得する．

一方，テスト事例 x の予測クラス $\hat{y}$ は，決定集合 R^* を用いて以下の様に算出される．

$$\hat{y} = \begin{cases} c & (P = \{c\}) \\ \text{By tie-breaking rule} & (|P| > 1) \\ \text{By default rule} & (P = \emptyset) \end{cases}$$

where $P = \{c \mid (s, c) \in R^*, s \subseteq x\}$.

すなわち，R^* 中の各ルールを独立に考え，x に適用できるすべてのルールが同じクラス c を予測した場合（$P = \{c\}$）はその値 c を採用する．また，予測クラスに衝突がある場合（$|P| > 1$）や適用できるルールが存在しない場合（$P = \emptyset$）は，それぞれ利用者が設定する衝突解消規則とデフォルト規則を用いてクラスを決定する．

以下では，各評価関数 f_i について説明する．

モデルの簡潔性：　関数 f_1 および f_2 はモデル R の簡潔性を評価する関数であり，それぞれルール数が少ない場合および各ルールが少ない条件数で構築されている場合に高い評価値が与えられる．

$$f_1(D, R) = |S| - |R|$$
$$f_2(D, R) = |S| \cdot \max\{|s'| \mid (s', c') \in S\} - \sum_{(s,c) \in R} |s|$$

モデルの有意義性：　関数 f_3 および f_4 はモデルの有意義性を評価する関数であり，各ルール間で被覆する訓練例集合の重複が小さいほど高い評価値を与える．

$$f_3(D, R) = |D| \cdot |S|^2 - \sum_{(s_i,c_i),(s_j,c_j) \in R, i<j, c_i=c_j} |\{\langle x, y \rangle \in D \mid s_i \subseteq x, s_j \subseteq x\}|$$

$$f_4(D, R) = |D| \cdot |S|^2 - \sum_{(s_i,c_i),(s_j,c_j) \in R, i<j, c_i \neq c_j} |\{\langle x, y \rangle \in D \mid s_i \subseteq x, s_j \subseteq x\}|$$

また関数 f_5 では，予測可能なクラス数という面からモデルの有意義性を評価する．

$$f_5(D, R) = |\{c \mid (s, c) \in R\}|$$

モデルの予測精度：　関数 f_6 および f_7 はモデルの精度を評価する関数であり，f_6 は適合性，f_7 は再現性の概念にそれぞれ相当する．

$$f_6(D, R) = |D| \cdot |S| - \sum_{(s,c) \in R} |\{\langle x, y \rangle \in D \mid s \subseteq x, y \neq c\}|$$

$$f_7(D, R) = |\{\langle x, y \rangle \in D \mid \exists (s, c) \in R \ s.t. \ s \subseteq x, y = c\}|$$

4　グラフ構造データへの拡張

IDS のグラフ構造データへの拡張として，(1) グラフ構造データからのクラス相関ルールの抽出，および (2) グラフ構造に着目した評価関数の導入を行う．

4.1　クラス相関ルールの獲得

グラフ $g = (V_g, E_g, l_g)$ を，頂点集合 V_g と辺集合 $E_g \subseteq V_g \times V_g$，ラベル関数 $l_g : V_g \cup E_g \to \mathcal{L}$（$\mathcal{L}$ はラベルの全体集合）の 3 項組で表現する．グラフ g_i とそのクラスラベル y_i の対 $t_i = \langle g_i, y_i \rangle$ をトランザクションとするデータベース $G = \{t_1, \cdots, t_{|G|}\}$ からクラス相関ルールを抽出することを考える．

部分グラフ p の G における支持度を

$$\mathrm{S}(p, G) = |\{\langle g, y \rangle \in G \mid p \sqsubseteq g\}| / |G|$$

と定義する．ここで $g_i \sqsubseteq g_j$ は，g_i が g_j の部分グラフであることを表す．支持度に関する閾値 σ $(0 < \sigma \leq 1)$ に

対し，条件 $\mathrm{S}(p,G) \geq \sigma \wedge \forall q \sqsubseteq p\,(\mathrm{S}(q,G) > \mathrm{S}(p,G))$ を満たす部分グラフを頻出フリー部分グラフ [11] と呼び，その全体集合を $\mathcal{F}_G^\sigma$ と表記する．

以上の準備の下，$\mathcal{F}_G^\sigma$ を用いて G から頻出フリー部分グラフ集合をアイテムとするトランザクションデータベース

$$D_G^\sigma = \{\langle\{p \in \mathcal{F}_G^\sigma \mid p \sqsubseteq g_i\}, y_i\rangle \mid \langle g_i, y_i\rangle \in G\}$$

を構築し，アイテム集合に対する一般的な手法を用いてクラス相関ルール $r=(s,c) \in S_G^\sigma \subseteq 2^{\mathcal{F}_\mathcal{G}^\sigma} \times C$ の集合 S_G^σ を導出する．なお，ルール $r=(s,c)$ の条件部が頻出フリー部分グラフの集合である（$s \subseteq \mathcal{F}_G^\sigma$）ことに着目し，条件 $\exists g_i, g_j \in s\ s.t.\ g_i \sqsubseteq g_j$，すなわち部分グラフの関係にあるグラフ同士を含むルールを S_G^σ から排除し，冗長性を排除する．

以降では簡略化のため，D_G^σ を D と，また S_G^σ を SS とそれぞれ同一視する．

4.2 グラフ構造に着目した評価関数の導入

ルール条件部が部分グラフパターン集合であることに着目し，評価関数 $F(D,R)$ を拡張する．具体的には，簡潔性の観点から新たに 3 つの評価関数 $f_i\ (i=\{8,9,10\})$ を導入し，パラメタ $\lambda_i\,(>0)$ を伴う非単調な劣モジュラ関数

$$F_G(D,R) = F(D,R) + \sum_{i=8}^{10} \lambda_i f_i(D,R)$$

を最適化基準として採用する．

以下，各評価関数を導入する．

グラフサイズ： 関数 f_8 は，ルールに含まれるグラフサイズ（頂点数＋辺数）の総計に基づき値を算出する．なお，サイズが小さいほど高い評価を与えるものとする．

$$f_8(D,R) = \sum_{(s,c)\in R} \left(\max_{(s',c')\in S} \{\mathrm{size}(s')\} - \mathrm{size}(r) \right)$$
$$\text{where } \mathrm{size}(s) = \textstyle\sum_{(V,E,l)\in s} |V| + |E|$$

頂点・辺のラベル種： 関数 f_9 は，グラフ中の頂点ラベル，辺ラベルに着目した評価関数であり，総種類数が多いほど高い評価を与える．

$$f_9(D,R) = \\ |\{l(o) \mid o \in V \cup E,\ (V,E,l) \in s, (s,c) \in R\}|$$

表 1: データセットに関する統計量

	クラス Y			クラス N		
	$\lvert V\rvert$	$\lvert E\rvert$	$\lvert\lambda\rvert$	$\lvert V\rvert$	$\lvert E\rvert$	$\lvert\lambda\rvert$
平均	25.49	26.63	9.84	24.31	25.57	9.34
分散	3.92	4.56	0.01	6.78	7.55	0.02
最小値	14.00	13.00	8.00	2.00	1.00	1.00
Q1	21.00	21.00	9.00	20.00	20.00	9.00
中央値	24.00	26.00	10.00	23.00	24.00	9.00
Q3	28.50	31.00	10.00	27.00	29.50	10.00
最大値	49.00	52.00	11.00	56.00	59.00	11.00

$|V|$ は頂点数，$|E|$ は辺数，$|\lambda|$ はラベル種数を表す

ルール内非類似度： 関数 f_{10} はグラフの非類似度に着目した評価関数であり，各ルール条件部が互いに非類似な部分グラフで構成される場合に高い評価を与える．なお今回は，各ルール $r=(s,c)$ の条件部 s に含まれる部分グラフ g_i, g_j 間の非類似度 $\mathrm{ged}(g_i,g_j)$ をグラフ編集距離を用いて算出し，その平均をルール r における非類似度としている．

$$f_{10}(D,R) = \sum_{(s,c)\in R} \frac{1}{{}_{|s|}\mathrm{C}_2} \sum_{g_i,g_j \in s, i<j} \mathrm{ged}(g_i,g_j)$$

5 評価実験

5.1 データセットと実験設定

提案手法の有効性を確認するため，株式会社 LIFULL が国立情報学研究所の協力により研究目的で提供している「LIFULL HOME’S データセット [1]」を基に構築されたグラフデータセット [8] を用いた実験を行った．

データセットは，東京 23 区内のマンションの各間取りをグラフ表現したものであり，専有面積や最寄り駅までの徒歩距離，近隣相場等を用いた推定賃料と実際の賃料との差から，クラス属性が付与されている．クラス属性は 2 値 (Y,N) であり，データ数は各クラス 240 件，合計 480 件である．表 1 にデータセットの概要を示す．なお，辺ラベルは存在せず，頂点のラベル種数は 11 である．

頻出部分グラフマイナー gSpan[12] と適切な後処理を用い，7,616 件の頻出フリー部分グラフを抽出した後，クラス相関ルール（最小支持度 0.3，最小確信度 0.5）の全体集合 $S^{all} = S^Y \cup S^N$ を導出した．ここで S^C はクラス値が C であるクラス相関ルールの集合であり，それぞれのルール数は，$|S^{all}| = 68,607$，$|S^Y| = 67,706$，$|S^N| = 901$ である．表 2 に得られたルールに関する統計量を示す．実験では，S^Y と S^N からそれぞれ 500 件

[1] https://www.nii.ac.jp/dsc/idr/lifull/homes.html

表 2: クラス相関ルールに関する統計量

	$\lvert s\rvert$	sup	conf	size	$\lvert\lambda\rvert$	$\overline{\text{ged}}$
			クラス Y			
平均	2.95	53.91	59.59	23.21	6.00	0.52
分散	3.61	0.03	4.38	0.01	5.90	3.41
最小値	1.00	46.71	50.00	2.00	1.00	0.00
Q1	3.00	49.13	57.14	18.00	5.00	0.37
中央値	3.00	51.90	59.87	24.00	6.00	0.50
Q3	3.00	57.09	62.24	28.00	7.00	0.67
最大値	3.00	96.89	65.04	44.00	8.00	1.00
			クラス N			
平均	2.87	75.08	51.37	14.01	4.28	0.64
分散	2.00	0.61	0.01	0.16	0.01	1.44
最小値	1.00	58.13	50.00	2.00	1.00	0.00
Q1	3.00	64.70	50.48	12.00	4.00	0.56
中央値	3.00	72.66	51.09	14.00	4.00	0.67
Q3	3.00	83.05	52.14	16.00	5.00	0.75
最大値	3.00	98.62	53.94	26.00	7.00	1.00

$|s|$ はルール本体長，sup は支持度，conf は確信度，size はグラフサイズ，$|\lambda|$ はラベル種数，$\overline{\text{ged}}$ は平均グラフ編集距離を表す.

のルールをランダムに選択することでクラス相関ルール集合 $S_i^{all} = S_i^Y \cup S_i^N$ $(i \in \{1, \cdots, 10\})$ を構築し，評価関数 $F(D, S_i^{all})$ に従い最適部分集合 $R_i^* \subseteq S_i^{all}$ を抽出するという操作を 10 回繰り返した．また，表 3 に示す各評価基準それぞれに関し，R_i^* $(i \in \{1, \cdots 10\})$ の平均値を求めることで評価を行った．

ここで各評価基準の概要について説明する．#_Rules と Avg_Length は，モデルの簡潔性に関する評価基準であり，それぞれ f_1 と f_2 に対応し，値が小さい方が優れていると判断する．一方，モデルの有意義性に関しては Frac_Overlap と Frac_Classes を用いて評価する．Frac_Overlap は f_3, f_4 に対応する．値域は $0 \sim 1$ であり，値が小さい方が優れていると判断する．Frac_Classes はクラスに関する網羅性を表し f_5 に対応する．今回は 2 値分類を対象としているので値域は $0.5 \sim 1$ となり，1 に近いほど良いと考える．モデルの精度に関しては，Frac_Uncovered と AUC を用いて評価する．Frac_Uncovered は説明されない事例の割合を表し f_7 に対応する．値域は $0 \sim 1$ であり，小さい方が優れていると考える．

評価基準 Graph_Size，Node_Type，Graph_Dist はグラフ構造に焦点を当てた評価基準であり，それぞれ f_8，f_9，f_{10} に対応する．Graph_Size はグラフの大きさの総計であり，値が小さい方が優れていると判断する．一方，ノード種とルール内非類似度に相当する Node_Type と Graph_Dist は値が大きい方が優れていると考える．

表 3: 評価基準

$$
\begin{aligned}
\#\_\mathrm{Rules}(R) &= |R| \\
\mathrm{Avg\_Length}(R) &= \tfrac{1}{|R|} \textstyle\sum_{(s,c)\in R} |s| \\
\mathrm{Frac\_Overlap}(R) &= \\
\frac{|D|}{{}_{|s|}\mathrm{C}_2} & \sum_{(s_i,c_i),(s_j,c_j)\in R, i<j} |\{\langle x, y\rangle \in D \mid s_i \subseteq x, s_j \subseteq x\}| \\
\mathrm{Frac\_Classes}(R) &= |\{c \mid (s,c) \in R\}|/|C| \\
\mathrm{Frac\_Uncovered}(R) &= \\
& 1 - \left|\cup_{(s,c)\in R}\{\langle x,y\rangle \in D \mid s \subseteq t\}\right|/|D| \\
\mathrm{Graph\_Size}(R) &= \textstyle\sum_{(s,c)\in R} \mathrm{size}(s) \\
\mathrm{Node\_Type}(R) &= f_9(D,R) \\
\mathrm{Graph\_Dist}(R) &= f_{10}(D,R)
\end{aligned}
$$

表 4: 実験結果

	F_G	$F_G^{\{8\}}$	$F_G^{\{9\}}$	$F_G^{\{10\}}$	$F_G^{\{8,9,10\}}$
AUC	0.58	0.52	0.58	0.58	0.52
#_Rules	4.20	2.00	4.20	4.20	2.00
Ave_Length	3.00	2.95	3.00	3.00	2.95
Frac_Overlap	0.00	0.35	0.00	0.00	0.35
Frac_Classes	1.00	1.00	1.00	1.00	1.00
Frac_Uncovered	0.11	0.22	0.11	0.11	0.22
Graph_Size	10.94	12.25	10.88	10.94	11.95
Node_Type	9.40	7.90	9.30	9.40	7.80
Graph_Dist	0.50	0.27	0.50	0.50	0.26

5.2 結果と考察

実験結果を表 4 に示す．なお表中の F_G^M は，R^* の選択基準として評価関数

$$F_G^M(D,R) = \sum_{i\in\{1,\cdots,10\}\setminus M} \lambda_i f_i(D,R)$$

すなわち M に含まれる評価関数を取り除いた場合を採用したことを表す．また $F_G^{\{8,9,10\}}(D,R) = F(D,R)$ は IDS オリジナルの評価関数に相当する．

提案した評価関数 F_G と IDS における評価関数 $F_G^{8,9,10}$ の結果を比較すると，精度（AUC と Frac_Uncovered）や有意義性（Frac_Overlap）の面で F_G の方が優れていることが確認できた．一方で，ルール数（#_Rules）は $F_G^{8,9,10}$ の方が値が小さく，簡潔さの面では一部劣る結果となったが，実際の値は#_Rules = 4.20 と十分に小さく，精度（AUC）とのバランスを考えると必ずしも大きな問題であるとは言えない．このことから，新たに導入した評価関数 $f_8 \sim f_{10}$ は概ね有効に働いていることと結論付けることができる．

次に $f_8 \sim f_{10}$ のそれぞれの効果を確認するため，$F_G^8 \sim F_G^{10}$ の結果について考察する．表 4 より，F_G^8 と $F_G^{8,9,10}$ の結果が類似していること，および F_G^9, F_G^{10}

と F_G の結果が類似していることがそれぞれ確認できる．F_G^8 はグラフサイズに関する評価関数 f_8 を排除した場合であり，F_G^8 の結果が $F_G^{8,9,10}$ と類似するということから，グラフサイズが精度と有意義性の向上に強く寄与していることが示唆される．一方で，F_G^9 と F_G^{10} の結果が F_G の結果と類似するということから，f_9（頂点・辺のラベル種）と f_{10}（ルール内非類似度）は，共に簡潔さに関する評価関数ではあるが，簡潔さの改善に寄与しないばかりか，精度・有意義性についても大きな影響を持たないことが伺える．以上の結果より，今回利用したデータセットに関して言えば，精度を保ったまま解釈容易なルールを導出するためにはグラフサイズを考慮することが有効であると結論付けることができる．

6 まとめ

本研究では，分類規則抽出に関する代表的な解釈性研究の一つである解釈可能決定集合を対象に，その適用範囲をグラフ構造データへと拡張する枠組みを提案した．また実データを用いた実験により，その可能性を示した．

今後の課題としては，更なる評価関数の開発と，他の実データを用いた評価があげられる．また，解釈性という観点から，人間による主観評価の実施も検討している．

謝辞： 本研究では，株式会社 LIFULL が国立情報学研究所の協力により研究目的で提供している LIFULL HOME'S データセットを利用した．なお本研究の一部は，JSPS 科研費 17K00315 の助成を受けたものです．

参考文献

[1] A. B. Arrieta et al.: Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI, *Information Fusion*, Vol.58, pp.82–115, 2020.

[2] 原 聡：私のブックマーク「機械学習における解釈性」，人工知能，Vol.33, No.3, pp.366–369, 2018.

[3] H. Lakkaraju, S. H Bach, and J. Leskovec : Interpretable Decision Sets: A Joint Framework for Description and Prediction, *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1675–1684, 2016.

[4] H. Cheng, X. Yan, and J. Han : Mining Graph Patterns, In C. Aggarwal, and H. Wang (eds) *Managing and Mining Graph Data*, pp.365–392, Springer, 2010.

[5] B. Liu, W. Hsu, and Y. Ma : Integrating Classification and Association Rule Mining, *Proc. of the 4th International Conference on Knowledge Discovery and Data Mining*, pp.80–86, 1998.

[6] H. Deng : Interpreting Tree Ensembles with inTrees, *International Journal of Data Science and Analytics*, Issue 4/2019, 2018.

[7] S. Hara and K. Hayashi : Making Tree Ensembles Interpretable: A Bayesian Model Selection Approach, *Proc. of the 21st International Conference on Artificial Intelligence and Statistics*, pp.77-85, 2018.

[8] T. Ozaki : Extraction of Characteristic Subgraph Patterns with Support Threshold from Databases of Floor Plans, *Proc. of the 2019 Seventh International Symposium on Computing and Networking*, pp.197–203, 2019.

[9] H. Lakkaraju, E. Kamar, R. Caruana, and J. Leskovec : Faithful and Customizable Explanations of Black Box Models, *Proc. of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp.131–138, 2019.

[10] U. Feige, V. S. Mirrokni, and J. Vondr´k : Maximizing non-monotone submodular functions, *SIAM Journal on Computing*, Vol.40, No.4, pp.1133–1153, 2011.

[11] Z. Zeng, J. Wang, J. Zhang, and L. Zhou : FOGGER : An Algorithm for Graph Generator Discovery, *Proc. of 12th International Conference on Extending Database Technology*, pp.517–528, 2009.

[12] X. Yan and J. Han : gSpan : Graph-based Substructure Pattern Mining, *Proc. of the 2002 IEEE International Conference on Data Mining*, pp.721-724, 2002.