部分グラフとその出現数を用いた グラフ構造データからの拡張相関ルールの獲得

Extraction of advanced association rules in graph databases based on subgraph patterns and their number of occurrences

尾崎 知伸 *
Tomonobu Ozaki

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

Abstract: Frequent pattern enumeration and association rule discovery are recognized as fundamental techniques in data mining. A wide variety of characteristic patterns and rules have been proposed to find interesting knowledge hidden in the transaction databases. However, compared with association mining in transaction databases, variation of patterns and rules are still limited for structured databases. In this paper, to extend the applicability of structured data mining, two advanced association rules, indirect association rules and association action rules, are proposed. To obtain the advanced rules, we first convert a given graph database into transaction one in which an item consists of a pair of frequent pattern and its discretized support value. We then perform pattern mining and rule discovery considering the subgraph relationship among attributes.

1 はじめに

系列や木，グラフ構造等に代表される構造化データを対象としたデータ分析技術は構造データマイニング [1] と総称され，頻出パターンの列挙 [2, 3, 4] を中心に，様々な側面からの研究が行われている．例えば，飽和パターン [5] やフリーパタン [6] 等の頻出パターンの圧縮表現の獲得や，強い共起関係を持つ部分グラフ集合である相互関連部分グラフ集合 [7] の列挙，高い排他性を有する部分グラフ集合である条件付対比部分グラフ [8] の列挙等は，代表的な研究例である．加えて，対象とするデータ構造の拡張として，グラフ系列 [9] や階層的構造データ [10] を対象とした頻出パターン列挙技術も提案されている．しかし，アイテム集合トランザクションデータベースを対象とした従来のパターン発見技術 [11] と比較した場合，現在の部分構造パターン発見技術で導出可能なパターンやそれらの関係性を表すルールの種類は限定的であり，より現実的な応用や深い分析を考えた場合，構造化データを対象としたより柔軟かつ多様な技術が必要とされる．

本研究では，構造データマイニングの高度化を目的とし，トランザクションデータベースを対象とした 2 種の拡張相関ルールをグラフデータベース用へと拡張することを提案する．第一のルールは間接相関ルール [12] であり，ある条件下での対比関係を表す相関ルール対である．また第二のルールは相関行動ルール [13] であり，ある条件下での属性値間の変化に関する関連性を表す相関ルール対である．本研究では，両拡張相関ルールの基本思想をグラフデータベースへと適用するため，まず頻出パターンを用い，分析対象であるグラフデータベースを，パターンとその支持度の対を最小単位（アイテム）とするトランザクションデータベースへと変換する．その上で，既存技術を素直に拡張すると共に，パターン間の包含関係を考慮することで,（グラフ構造パターンと支持度の対を基本要素とする）頻出パターンの列挙，相関ルールの導出，拡張相関ルールの獲得へと繋げる．

本論文の構成は以下の通りである．2 章では準備として頻出部分グラフマイニングに関する定義を与える．次いで 3 章で，グラフ構造パターンと支持度の対をアイテムとする頻出パターンを導入する．また，グラフデータに対する間接相関ルールを 4 章で，同じく相関行動ルールを 5 章でそれぞれ提案する．最後に 6 章でまとめを行い，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　頻出部分グラフマイニング

ラベルの全体集合を $\mathcal{L}$ とする．また，ラベル付きグラフ $g = (V_g, E_g, \lambda_g)$ を，頂点集合 V_g，辺集合 $E_g \subseteq V_g \times V_g$，ラベル関数 $\lambda_g : V_g \cup E_g \to \mathcal{L}$ の 3 項組で表す．2 つのラベル付きグラフ p, g に関し，以下の条件

1. $\forall u \in V_p\, [\lambda_p(u) = \lambda_g(f(u))]$
2. $\forall (u, v) \in E_p$
$\exists (f(u), f(v)) \in E_g\ s.t.\ \lambda_p(u, v) = \lambda_g(f(u), f(v))$

を満足する頂点間の写像関数 $f : V_p \to V_g$ が存在するとき，p を g の部分グラフと呼び，$p \sqsubseteq g$ と表記する．

グラフデータベース $G = \{g_1, \cdots g_n\}$ に対し，条件 $p \sqsubseteq g \in G$ を満たすグラフ p を部分グラフパターンと呼ぶ．G における部分グラフパターン p の支持度を

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

と定義する．すなわち p の支持度とは，部分グラフとして p を持つグラフの割合である．

支持度に関する閾値 $\sigma\,(0 < \sigma \leq 1)$ に対し，条件 $S(p, G) \geq \sigma$ を満足する部分グラフパターン p を G における頻出部分グラフパターンと呼ぶ．また，支持度を基準とする誤差を考慮した頻出部分グラフパターンの同値類における極小元として，誤差に関するパラメタ $\delta \geq 0$ に対し，条件

$$\forall q \sqsubseteq p\, (\, S(q, G) > (1 + \delta) \cdot S(p, G)\,)$$

を満たす頻出部分グラフパターン p を，頻出 δ フリー部分グラフパターンと呼ぶ．なお，G に対する頻出 δ フリー部分グラフパターンの全体集合

$$\mathcal{F}_G^{\sigma,\delta} = \left\{ p \,\middle|\, \begin{array}{l} S(p, G) \geq \sigma \,\wedge \\ \forall q \sqsubseteq p\, (\, S(q, G) > (1 + \delta) \cdot S(p, G)\,) \end{array} \right\}$$

は，頻出部分グラフパターンの全体集合に対する一つの圧縮表現に相当する．以降簡略化のため，頻出 δ フリー部分グラフパターンを頻出パターンと呼ぶ．

3　頻出グラフ属性集合

本章ではまず，頻出パターンを属性として利用することで，グラフデータベースをトランザクション化することを考える．その後，トランザクション化されたデータベースに対する頻出アイテム集合発見について議論する．

3.1　単一グラフに対するパターンの支持度

頻出 δ フリー部分グラフパターン集合 $\mathcal{F}_G^{\sigma,\delta}$ を用いたグラフデータベース G のトランザクション化においては，各グラフ $g \in G$ に対する頻出パターン $p \in \mathcal{F}_G^{\sigma,\delta}$ の支持度を利用する．

一般に，単一グラフにおけるパターンの支持度は自明ではなく，これまで種々の定義が考案されている．本研究では文献 [14, 15] に従い，グラフデータ $g \in G$ における属性 $p \in \mathcal{F}_G^{\sigma,\delta}$ の出現，すなわち p と同型である g の部分グラフ $o = (V_o, E_o, \lambda_o)$ の全体集合

$$O(p, g) = \{o \mid o \sqsubseteq g, p \sqsubseteq o, o \sqsubseteq p\}$$

に基づく各種の支持度を利用する．

①極大独立集合に基づく支持度 [16]

集合 $MIS(p, g)$ を，各出現 $o \in O(p, g)$ を頂点とし，（g における）頂点を共有する出現同士を辺で結ぶことで得られるグラフ

$$\big(O(p, g),\, \{(o_j, o_k) \mid o_j, o_k \in O(p, g), V_{o_j} \cap V_{o_k} \neq \emptyset\},\, \lambda \big)$$

の極大独立集合とする．このとき，極大独立集合に基づく支持度 $S_{MIS}(p, g)$ は，$MIS(p, g)$ を用い，

$$S_{MIS}(p, g) = |MIS(p, g)|,$$

すなわち頂点を共有せずに g 中に配置することのできる出現 $o \in O(p, g)$ の最大数と定義される．

②最制限頂点に基づく支持度 [17]

頂点集合 V_p, V_g と出現 $o = (V_o, E_o, \lambda_o) \in O(p, g)$ に対し，関数 $\varphi_o : V_p \to V_g$ を条件

1. $\forall v \in V_p \Rightarrow \lambda_p(v) = \lambda_g(\varphi(v))$
2. $\forall (u, v) \in E_p \Rightarrow (\varphi(u), \varphi(v)) \in E_g$

を満たす頂点間の写像関数とする．このとき最制限頂点に基づく支持度 $S_{MRN}(p, g)$ は，p 中の頂点 $v \in V_p$ の写像先集合に対する最小要素数であり，

$$S_{MRN}(p, g) = \min_{v \in V_p} |\{\varphi_o(v) \mid o \in O(p, g)\}|$$

と定義される．

支持度 $S_{MIS}(p, g)$ では，出現間の重複を許容しないという制約の元で，出現を単位としたカウントを行っている．これに対し支持度 $S_{MRN}(p, g)$ では，出現単位ではなく，p 中の頂点の単位で，写像先の重複を考慮した形でのカウントを行っている．

③出現数に基づく支持度 [14, 15]

支持度 $S_{NOO}(p,g)$ は，出現 $o \in O(p,g)$ における辺集合 E_o の種類数，すなわち

$$S_{NOO}(p,g) = |\{E_o \mid (V_o,E_o,\lambda_o) \in O(p,g)\}|$$

と定義される．定義が示す通り，$S_{NOO}(p,g)$ では完全な重複以外を許容している．しかし，完全な重複以外を許容する結果として，例えば一ヶ所だけが異なるような出現を多数持つ部分グラフに対して不当に大きな値を与えてしまう可能性がある点に注意が必要である．

④頂点数の比に基づく支持度 [14, 15]

支持度 $S_{ROV}(p,g)$ は，出現の重複に関する正規化処理を採用した支持度であり，

$$S_{ROV}(p,g) = |\{v \in V_o \mid (V_o,E_o,\lambda_o) \in O(p,g)\}| \,/\, |V_p|$$

と定義される．正規化処理，すなわち出現に含まれる総頂点数を部分グラフの頂点数で除することで，支持度 $S_{NOO}(p,g)$ が持つ問題の回避を意図している．

3.2 グラフ属性上の頻出アイテム集合

グラフデータベース G とそこから得られる頻出パターン集合 $\mathcal{F}_G^{\sigma,\delta}$ に関し，属性を頻出パターン，その属性値を離散化された支持度とする属性・値対をアイテムとするトランザクションデータベースを

$$\begin{aligned} T(G) &= \{t(g) \mid g \in G, t(g) \neq \emptyset\} \quad \text{where} \\ t(g) &= \left\{\langle p, d_p(\, S_x(p,g)\,)\,\rangle \mid p \in \mathcal{F}_G^{\sigma,\delta}, S_x(p,g) > 0\right\} \end{aligned}$$

と定義する．ここで d_p は，グラフ g に対する頻出パターン p の支持度 $S_x(p,g)$ ($x \in \{MIS, MRN, NOO, ROV\}$) を離散化する関数である．また各グラフ g 毎に，g が含む頻出パターンのみを対象に属性･値対 $\langle p, d_p(\, S_x(p,g)\,)\,\rangle$ を抽出している．

本研究では，トランザクションデータベース $T(G)$ に対する特徴的なパターンの一つとして，属性である頻出パターン間の包含関係を考慮した属性･値対の集合を考える．属性･値対の集合 $P = \{\langle p_1, v_1\rangle, \cdots, \langle p_m, v_m\rangle\}$ に対し，$T(G)$ におけるその支持度を

$$S(P, T(G)) = |\{t \in T(G) \mid P \subseteq t\}| \,/\, |T(G)|,$$

すなわち P を含むトランザクションの割合と定義する．一方，P に含まれる頻出パターンの集合を $AT(P) = \{p \mid \langle p, v\rangle \in P\}$ と表記し，包含関係を持つ頻出パターンを含まないという包含関係に関する制約を

$$c_{\sqsubseteq}(P) \equiv \forall p_i, p_{j\,(i \neq j)} \in AT(P)\ p_i \not\sqsubseteq p_j$$

DFS(P)
for each $\langle q, v\rangle$ s.t. $q \in \mathcal{F}_G^{\sigma,\delta}$, $v \in D_q$, $q \underset{lex}{>} tail(P)$
$Q := P \cup \{\langle q, v\rangle\}$
if $S(Q, T(G)) \geq \varsigma \wedge$
$\forall p \in AT(P)\,(p \not\sqsubseteq q \wedge q \not\sqsubseteq p)$ **then**
output(Q)
DFS(Q)
end if
end for each

図 1: 縦型探索による頻出グラフ属性集合の列挙

と表記する．以上の準備の下，支持度に関する閾値 ς ($0 < \varsigma \leq 1$) に対し，空でない属性・値対の集合 $P\,(\neq \emptyset)$ が条件 $S(P, T(G)) \geq \varsigma \wedge c_{\sqsubseteq}(P)$ を満たすとき，P を G に対する頻出属性・値対集合パターンと呼ぶ．以降簡略化のため，頻出属性・値対集合パターンを頻出グラフ属性集合と呼ぶ．また，G に対する頻出グラフ属性集合の全体集合を $\mathcal{I}_G^{\varsigma}$ と表記する．

ところで，属性・値対の集合は，支持度と包含関係に関する制約に対し，逆単調性を満たす．より形式的には，2 つの属性・値対集合 P と Q に対し，

$$P \subseteq Q \rightarrow \begin{pmatrix} S(P, T(G)) \geq S(Q, T(G)) \\ \neg c_{\sqsubseteq}(P) \rightarrow \neg c_{\sqsubseteq}(Q) \end{pmatrix}$$

が成り立つ．このことから，apriori アルゴリズム [18] 等の従来の頻出アイテム集合列挙アルゴリズムを少し拡張するだけで，頻出グラフ属性集合の列挙が達成可能であることが分かる．例として，縦型探索（バックトラック法）に基づく列挙アルゴリズムを図 1 に示す．なお図 1 において，D_q は頻出パターン q に対する属性値の離散化結果の値域を，$tail(P)$ は P の末尾（最後に追加された要素）を表す．また関係 $\underset{lex}{>}$ は，グラフの文字列表現である DFS コード [4] などを用いた頻出パターン間の前後関係を表す．

4 グラフ属性間接相関ルール

本章ではまず，グラフ属性集合を基にした属性間の関係を表すルール，すなわち相関ルールを導入する．次いで，間接相関ルール（Indirect Association Rule）[12] の考えを適用し，グラフデータベースにおける，ある条件下での対比関係を表すルール対を提案，考察する．

4.1 グラフ属性上の相関ルール

グラフデータベース G と，2 つの頻出グラフ属性集合 $P, Q \in \mathcal{I}_G^{\varsigma}, P \cap Q = \emptyset$ に対し，「G においては，P を

含むトランザクションは Q も含む傾向が高い」こと表すルール $P \to Q$ を考える．従来の相関ルール発見技術に倣い，ルール $P \to Q$ の支持度 (S) と確信度 (C), リフト値 (L) を以下の様に定義する．また，確信度とリフト値を合わせて依存度と呼び，D と表記する．

$$\begin{aligned}
S(P \to Q, T(G)) &= S(P \cup Q, T(G)) \\
C(P \to Q, T(G)) &= S(P \to Q, T(G)) \,/\, S(P, T(G)) \\
L(P \to Q, T(G)) &= \frac{S(P \to Q, T(G))}{S(P, T(G)) \cdot S(Q, T(G))}
\end{aligned}$$

支持度に関する閾値 $\varsigma\,(0 < \varsigma \leq 1)$ に加え，依存度に関する閾値 $\vartheta\,(0 < \vartheta)$ に対し，条件

$$S(P \to Q, T(G)) \geq \varsigma \wedge D(P \to Q, T(G)) \geq \vartheta$$

を満たすルール $P \to Q$ を，グラフ属性相関ルールと呼ぶ．また，グラフデータベース G に対するグラフ属性相関ルールの全体集合を $\mathcal{R}_G^{\varsigma,\vartheta}$ と表記する．

4.2　グラフ属性上の間接相関ルール

間接相関ルールとは，ある条件下での対比関係を表す単一要素集合を頭部として持つ 2 つの相関ルールの対である．ここではアイテム集合を対象とした間接相関ルールの素直な拡張として，グラフ属性相関ルールに基づく間接相関ルールを導入する．

ルール本体部を共有する 2 つのグラフ属性相関ルール

$$R_p = M \to \{\langle p, v_p \rangle\}, R_q = M \to \{\langle q, v_q \rangle\} \quad (R_p, R_q \in \mathcal{R}_G^{\varsigma,\vartheta}, p \neq q)$$

を考える．R_p, R_q が，支持度の最大値に関する閾値 $\varsigma_{\max}\,(0 \leq \varsigma_{\max} \leq 1)$ に対して条件

$$S(\{\langle p, v_p \rangle, \langle q, v_q \rangle\}, T(G)) \leq \varsigma_{\max}$$

を満たすとき，その対 $\langle R_p, R_q \rangle$ をグラフデータベース G に対するグラフ属性間接相関ルールと呼ぶ．また特に，本体部と頭部を意識するため，グラフ属性間接相関ルール $\langle R_p, R_q \rangle$ を 3 項組 $\langle M, \langle p, v_p \rangle, \langle q, v_q \rangle \rangle$ の形式で表現することもある．

グラフ属性間接相関ルールは，その前提条件として $p \neq q$ を採用している．これは以下の理由によるものである．グラフ属性に対する属性値 (v_p, v_q) は頻出パターン p, q の支持度を離散化したものであり，$p = q$ の場合，共通する本体部 M とは無関係に頭部間の対比（排他）関係（$S(\{\langle p, v_p \rangle, \langle q, v_q \rangle\}, T(G)) = 0$）が成り立つ．この様な意味をなさないルール対を排除するため，$p \neq q$ を前提条件としている．

一方，ルール頭部のグラフ属性に着目すると，グラフ属性間接相関ルール $\langle M, \langle p, v_p \rangle, \langle q, v_q \rangle \rangle$ を，グラフ属性間に包含関係がある場合（$p \sqsubseteq q$）とない場合（$p \not\sqsubseteq q$）の 2 種類に分類することが可能である．特に前者に関しては，包含関係があるにもかかわらず $\langle p, v_p \rangle$ と $\langle q, v_q \rangle$ 間の共起が非常に小さいという状況を捉えている点，その違いや原因を 2 つのグラフ p, q の差を中心に考察することが可能な点で，非常に興味深いルールであると考えられる．また後者に関しても，間接相関ルールが入れ替え可能，すなわち同様の役割を果たすアイテム対の発見を目的としていることを踏まえると，一つの役割に対して構造的に大きな違いがあるグラフ属性を併記し，より多くの情報を用いた役割の分析を支援するという点で有益であると考えられる．

5　グラフ属性相関行動ルール

相関行動ルール（Association Action Rule）[13] とは，ある条件下において帰結を望ましい状態に変化させるために必要とされる前提の変化，すなわち必要な行動を明示化したルールであり，相関ルールの対として与えられる．本章では，クラス属性付きのグラフデータベースを対象とし，属性値の変化に加え，属性間の包含関係を考慮した相関行動ルールを提案する．

5.1　グラフ属性上のクラス相関ルール

グラフデータベース G 中の各グラフ $g \in G$ に対し，そのクラス属性 y の値を $y(g)$ と表記する．なおクラス属性は，外部情報を利用して付与することに加え，G における部分グラフパターンの有無や，その離散化支持度を利用することも考えられる．

G におけるクラスを予測するルールとして，頻出グラフ属性集合 $P \in \mathcal{I}_G^{\varsigma}$ を前提，クラス属性を帰結とするルール $P \to \langle y, v_y \rangle$ を考える．このときルールの支持度を

$$S(P \to \langle y, v_y \rangle, T(G)) = \frac{\left| \left\{ g \;\middle|\; \begin{array}{l} g \in G, y(g) = v_y, \\ P \subseteq t(g) \in T(G) \end{array} \right\} \right|}{|T(G)|}$$

確信度を

$$C(P \to \langle y, v_y \rangle, T(G)) = \frac{S(P \to \langle y, v_y \rangle, T(G))}{S(P, T(G))}$$

と定義する．以上の準備の下，2 つの閾値 $\varsigma\,(0 < \varsigma \leq 1)$ と $\vartheta\,(0 < \vartheta)$ に対して条件

$$S(P \to \langle y, v_y \rangle, T(G)) \geq \varsigma \wedge C(P \to \langle y, v_y \rangle, T(G)) \geq \vartheta$$

を満たすルール $P \to \langle y, v_y \rangle$ をグラフ属性クラス相関ルールと呼ぶ．またグラフ属性クラス相関ルールの全体集合を $\mathcal{CR}_G^{\varsigma,\vartheta}$ と表記する．

5.2 グラフ属性上の相関行動ルール

クラス属性付きのグラフデータベース G に対し，ルール本体部の一部を共有する 2 つのグラフ属性クラス相関ルール

$$\begin{array}{lcl} R_p & = & M \cup \{\langle p, v_p \rangle\} \rightarrow \langle y, v_{y_p} \rangle \\ R_q & = & M \cup \{\langle q, v_q \rangle\} \rightarrow \langle y, v_{y_q} \rangle \\ \multicolumn{3}{l}{(R_p, R_q \in \mathcal{CR}_G^{\varsigma,\vartheta}, M \neq \emptyset, \langle p, v_p \rangle \neq \langle q, v_q \rangle, v_{y_p} \neq v_{y_q})} \end{array}$$

を考える．ルール対 $\langle R_p, R_q \rangle$ に対し，その支持度と確信度をそれぞれ

$$\begin{array}{lcl} S(\langle R_p, R_q \rangle, T(G)) & = & \min(S(R_p, T(G)), S(R_q, T(G)) \\ C(\langle R_p, R_q \rangle, T(G)) & = & C(R_p, T(G)) \cdot C(R_q, T(G)) \end{array}$$

と定義する．支持度に関する閾値 $\sigma\,(0 < \sigma \leq 1)$ と確信度に関する閾値 $\theta\,(0 < \theta \leq 1)$ に対し，ルール対 $\langle R_p, R_q \rangle$ が条件

$$S(\langle R_p, R_q \rangle, T(G)) \geq \sigma \ \wedge \ C(\langle R_p, R_q \rangle, T(G)) \geq \theta$$

を満たすとき，$\langle R_p, R_q \rangle$ をグラフ属性相関行動ルールと呼ぶ．グラフ属性相関行動ルール $\langle R_p, R_q \rangle$ は,「グラフ属性集合 M の条件下で属性 $\langle p, v_p \rangle$ を属性 $\langle q, v_q \rangle$ に（もしくは $\langle q, v_q \rangle$ を $\langle p, v_p \rangle$ に）変化させることでクラスが v_{y_p} から v_{y_q} に（もしくは v_{y_q} から v_{y_p} に）変わる傾向が見られる」ことを表しており，クラス値を変えるために必要とされるグラフ属性の変化を示していると解釈できる．

グラフ属性相関行動ルールは，必要とされるグラフ属性の変化，すなわち $\langle p, v_p \rangle$ と $\langle q, v_q \rangle$ の関係に基づき 3 種に大別することが可能である．第 1 種は条件 $p = q$ を満たすルール集合であり，前提条件である $\langle p, v_p \rangle \neq \langle q, v_q \rangle$ より，属性 p を共有し属性値が異なる（$v_p \neq v_q$）ルール対ということになる．これらのルールは,「属性値の変化がクラスの変化に繋がる」というトランザクション集合を前提としたオリジナルの相関行動ルール [13] の最も自然な拡張であり，その解釈も容易であると考えられる．第 2 種は条件 $p \sqsubseteq q \wedge p \neq q$ を満たすルール集合である．これらのルールは，包含関係という点で関連性の高いグラフ属性間の入れ替えを捉えており，属性値の変化に加えグラフ構造の差を中心にクラスを変化させる要因を分析可能という点で興味深いと言える．第 3 種は条件 $p \not\sqsubseteq q$ を満たすルール集合である．これらのルールは，グラフ構造の観点から大きな変化の必要性を示しており，目的（クラス値の変化）を達成するために備えるべき部分構造と排除すべき部分構造を示しているという点で有益であると考えられる．この様に，クラス値の変化に対して段階的なルールが獲得可能であるという点は，グラフ属性を利用する大きな特徴の一つであると考えられる．

6 まとめと今後の課題

本論文では，グラフ構造データを対象としたパターンマイニング研究の発展を目的とし，グラフ属性の出現数を考慮することでアイテム集合マイニングにおける諸パターン・ルールを拡張し，グラフ構造データに対する頻出パターン，相関ルール，間接相関ルール，相関行動ルールを提案した．現状，形式的な定義の導入と簡単な考察を終えた段階であり，実データを用いた評価・考察は今後の課題とする．また併せて，効用に基づくパターン発見技術 [19] の拡張など，グラフ構造データを対象とした更なるパターンの開発を継続することを予定している．

謝辞： 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] C. C. Aggarwal and H. Wang (eds.) : *Managing and Mining Graph Data*, Springer, 2010.

[2] A. Inokuchi, T. Washio, and H. Motoda : An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data, *Proc. of the 4th European Conference on Principles of Data Mining and Knowledge Discovery*, pp.13-23, 2000.

[3] M. Kuramochi and G. Karypis : Frequent Subgraph Discovery, *Proc. of the 2001 IEEE International Conference on Data Mining*, pp.313–320, 2001.

[4] X. Yan and J. Han : gSpan:Graph-based substructure pattern mining, *Proc. of the 2002 IEEE International Converence on Data Mining*, pp.721–724, 2002.

[5] X. Yan and J. Han : CloseGraph: Mining Closed Frequent Graph Patterns, *Proc. of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.286-295, 2003.

[6] Z. Zeng, J. Wang, J. Zhang, and L. Zhou : FOGGER: an algorithm for graph generator discovery, *Proc. of the 12th International Conference on Extending Database Technology Technology*, pp..517528, 2009.

[7] 尾崎 知伸, 大川 剛直 : グラフデータベースからの頻出相互関連部分グラフ集合の発見, 人工知能学会論文誌, Vol.23, No.6, pp.514-525, 2008.

[8] T. Ozaki and M. Etoh : Correlation and Contrast Link Formation Patterns in a Time Evolving Graph, *Workshops Proceedings of the 11th IEEE International Conference on Data Mining*, pp.1147-1154, 2011.

[9] A. Inokuchi and T. Washio : GTRACE: Mining Frequent Subsequences from Graph Sequences, *IEICE Transactions on Information and Systems*, Vol.E93.D, No.10, pp.2792-2804, 2010.

[10] 山本 翼, 三好 裕樹, 尾崎 知伸, 大川 剛直 : 複合構造グラフからの頻出強相関パターン発見, 情報処理学会論文誌 : データベース, Vol.2, No.3, pp.53-66, 2009.

[11] J. Han, H. Cheng, D. Xin, X. Yan, Frequent pattern mining: Current status and future directions, *Data Mining and Knowledge Discovery*, Vol.15, No.1, pp.55–86, 2007.

[12] P.-N. Tan, V. Kumar and J. Srivastava：Indirect Association: Mining Higher Order Dependencies in Data, *Proc. of the 4th European Conference on Principles of Data Mining and Knowledge Discovery*, pp.632–637, 2000.

[13] Z. W. Ras, A. Dardzinska, L.-S. Tsay, and H. Wasyluk. Association Action Rules. *Proc. of the 2018 IEEE International Conference on Data Mining Workshop*, pp.283–290, 2008.

[14] 長谷川 優也，尾崎 知伸：部屋配置とその出現数に着目した二段階賃料推定，人工知能学会 第21回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，pp.80–87, 2019.

[15] T. Ozaki : Extraction of Characteristic Subgraph Patterns with Support Threshold from Databases of Floor Plans The 2019 Seventh International Symposium on Computing and Networking, 2019 (to appear)

[16] M. Kuramochi and G. Karypis : Finding frequent patterns in a large sparse graph, *Data Mining and Knowledge Discovery*, Vol.11, No.3, pp.243–271, 2005.

[17] B. Bringmann and S. Nijssen : What is Frequent in a Single Graph? *Proc.* of the 5th International Workshop on Mining and Learning with Graphs, pp.1–4, 2007.

[18] R. Agrawal, T. Imielinski and A.Swami : Mining association rules between sets of items in large databases, *Proc. of the 1993 ACM-SIGMOD International Conference on Management of Data*, pp.207–216, 1993.

[19] W. Gan, J. C.-W. Lin, P. Fournier-Viger, H.-C. Chao, V. S. Tseng, and P. S. Yu : A Survey of Utility-Oriented Pattern Mining, *CORR*, abs/1805.10511, 2018.