

交通系 IC カードデータからの交通弱者の移動パターン抽出 Travel Pattern Extraction of Elderly and Disabled from Smart Card Data

細江 美欧[†] 桑野 将司[†]
Mio Hosoe Masashi Kuwano

1. はじめに

地方都市では、自動車を運転できる住民は時間的かつ空間的な制約なく自由に外出可能である。一方で、自動車を運転できない高齢者や障害者、子供といった交通弱者は公共交通による移動に頼らざるを得ず、移動制約が多い。交通弱者の生活の質の向上のためには、公共交通の再編による移動手段の確保は重要な課題である。

交通弱者のニーズに対応した公共交通の運行計画策定のためには、交通弱者を含む公共交通利用者の交通行動を把握する必要がある。そのために、従前よりアンケート調査の結果を用いて交通行動を把握する研究¹⁾²⁾が取り組まれてきた。しかしながら、そのような紙面調査ではある特定の日あるいは短期間の交通行動しか把握できず、日々変化する交通需要が把握できない。

一方で、近年の技術革新によって多種多様なビッグデータが入手可能となり、交通行動を把握するためのデータとして、ETCデータ³⁾やGPSデータ⁴⁾、プローブカーデータ⁵⁾などの交通系ビッグデータへの注目が高まっている。これら交通系ビッグデータのひとつに交通系ICカードの利用履歴データがある。交通系ICカードの利用履歴データにはカード保有者が何時にどの駅から乗車し、何時にどの駅で降車したのかといった移動軌跡に関する情報が継続的に収集されており、長期的な交通需要の変化が把握可能である。

嶋本ら⁶⁾は交通系ICカードデータに含まれる料金支払い形態からカード利用者を分類し、分類した利用者区分や時間帯による利用回数の変動を明らかにした。さらに、分散分析を用いて、特定の利用者区分に属す利用者の利用回数の変動に影響を与えるイベントが、他の利用者区分の利用者の行動にも影響を与えることを示した。野上ら⁷⁾はカードID、乗降車地点、乗降車時間帯から、乗継ぎと乗換えを考慮した分析を行い、特定の時間帯における移動量の多い乗車地点および降車地点の把握を行った。また、利用回数の多い乗換え地点を定量的に明らかにした。日下部ら⁸⁾は交通機関利用者が1日の間に利用する降車駅の組合せに着目し、バスケット分析を用いて、平日と休日で降車駅の組合せに違いがあることを明らかにした。また、降車駅の組合せの多くは降車人数が比較的上位の駅と下位の駅の組合せであったことを示した。

これら研究のように交通系ICカードの利用履歴データを交通行動データとして捉え、交通行動を把握することによって、交通政策のための新たな検討材料を得られることが期待される。しかしながら、前述の研究の多くは、交通系ICカードの利用履歴データに含まれる多数のデータ項目の中から例えば、利用者区分と利用回数、乗車駅と降車駅などの少数の項目にのみ着目し、その関連性を分析している。言い換えると、利用履歴データに含まれる複数のデータ項目、すなわち高次数間の関連性を同時に考慮できていない。交通系ICカードの利用履歴データには乗車駅、

降車駅、利用時間帯、利用曜日、利用者区分など複数の次数についてのデータが記録されており、これら次数は互いに影響を及ぼし合っていると考えられる。しかし、分析者があるデータ項目のみを抽出し分析を行った場合、考慮していない次数が交通行動に影響を及ぼしている可能性を議論することができず、どのデータ項目に着目するかによって異なる傾向が抽出される危険性がある。ビッグデータの有効かつ効率的な解析の観点からも、可能な限りデータの持つ情報を同時に取り扱い、そこからデータの特徴を把握することが望まれる。

そこで、本研究では複数の次数を同時に取り扱うことができる方法論としてグラフ研磨に着目する。グラフ研磨は、宇野ら⁹⁾、中原ら¹⁰⁾によって提案されたデータから共通した特徴を持つ中規模のグループを抽出するためのクラスタリング手法である。グラフ研磨では類似指標を用いてグラフからグループを抽出するため、入力データにおいては少ないサンプル数で表現される頂点であっても、頂点間の類似性が高ければグループとして抽出される。そのため、全体に占める割合が低くても規則性のあるパターンや数は少ないが類似性のあるグループを抽出できる可能性がある。つまり、交通弱者のようなサンプル数が少ないためにパターン抽出されにくい利用者に関する交通行動も把握可能となると考えられる。

本研究では、交通系ICカードデータに含まれる乗車曜日（以下、曜日）、乗車時間帯（以下、時間帯）、利用者区分、乗車駅、降車駅の組合せ、すなわち5次数の情報からカード利用者の移動パターンを抽出するために、グラフ研磨を応用した新たな方法論の提案を目的とする。具体的には、曜日、時間帯、利用者区分の3次数を組込んだ「属性頂点」と、乗車駅、降車駅の2次数を組込んだ「起終点（以下、OD (Origin-Destination) 頂点）」とのつながりの強さからグループ化し、類似した移動をする属性頂点のグループを抽出する。さらに、抽出された属性頂点のグループ毎に利用量の多い乗車駅と降車駅の組合せを把握する。

2. 使用データの概要と基礎集計結果

2.1 使用データの概要

本研究では、香川県において利用されている交通系ICカード「IruCa」の利用履歴データを分析対象とする。IruCaは高松琴平電気鉄道株式会社によって運営されている鉄道路線「ことでん」全線52駅、バス路線「ことでんバス」全路線96両、および県内バス会社路線に導入されているが、本研究では鉄道路線「ことでん」における利用履歴データのみを用いる。

ことでんには図1に示すように、琴平線、長尾線、志度線の3路線全52駅が存在する。「高松築港駅」、「片原町駅」は琴平線および長尾線の乗降車駅となっている。さらに、「瓦町駅」は3路線の結節点となっており、3路線すべてに乗り換えが可能である。本研究では、利用履歴データ

[†]鳥取大学 Tottori University

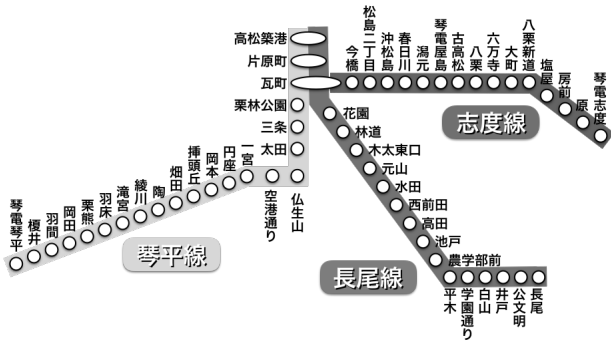


図1 ことでん路線図¹⁾

に蓄積された情報のうち、曜日 8 区分 (月曜日, 火曜日, ..., 日曜日, 祝日), 時間帯 20 区分 (5 時台, 6 時台, ..., 23 時台, 24 時台), 利用者 11 区分 (大人, 学生, 小人, 高齢者, 障害者, 定期大人通勤, 定期大人通学, 定期小人通勤, 定期小人通学, 定期障害者通勤, 定期障害者通学), 乗車駅 52 駅, 降車駅 52 駅の 5 次数を用いる。データの収集期間は, 2013 年 12 月 1 日から 2015 年 2 月 28 日までの 15 ヶ月間であり, この期間に収集されたデータ数は 9,033,748 件である。このうち, ことでんの運行時間帯である 5 時台から 24 時台に利用され, 移動所要時間が 60 秒以上のデータ 9,008,709 件を有効データとして用いる。

2.2 基礎集計結果

ことでんの利用状況を把握するために, 曜日, 時間帯, 利用者区別に平均延べ利用者数を算出した結果を, それぞれ図 2, 図 3, 図 4 に示す。

図 2 の曜日別利用状況の結果から, 月曜日から金曜日までの平日の平均延べ利用者数はほとんど変化がなく, その平均は約 24,794 件であり, 土曜日, 日曜日, 祝日の休日の平均延べ利用者数の平均約 9,026 件と比べて, 平均延べ利用者数は約 2.7 倍であることがわかった。これは通勤や通学などを目的とした利用者の平日利用の多さが影響していると考えられる。

図 3 の時間帯別利用状況の結果からは, 7 時台の平均延べ利用者数が最も多いことがわかる。また, 5 時台から 7 時台にかけて平均延べ利用者数が増加しており, これは通勤・通学を目的とした利用であると考えられる。8 時台から 11 時台にかけては平均延べ利用者数が減少しており, その後は 14 時台まで平均延べ利用者数の大きな変動は見られない。15 時台以降は平均延べ利用者数が再び増加しており, 特に 17 時台および 18 時台の平均延べ利用者数が多い。17 時台以降は帰宅を目的とした利用であると考えられる。

図 4 の利用者区別利用状況の結果からは, 大人, 定期大人通勤, 定期大人通学といった「大人」に関する平均延べ利用者数が多く, 特に定期大人通勤の平均延べ利用者数は全体の約 47% と多いことがわかった。一方で, 定期小人通勤, 定期障害者通勤, 定期障害者通学の平均延べ利用者数はそれぞれ約 0.00497%, 約 0.02696%, 約 0.00002% と少ないことが明らかとなった。

3. グラフ研磨を用いた移動パターン抽出方法

3.1 記法

V を頂点集合, E を辺集合とし, $E \subset [V]^2$ を満たす 2 つの

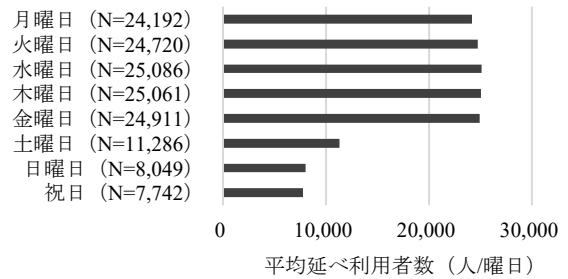


図2 曜日別利用状況

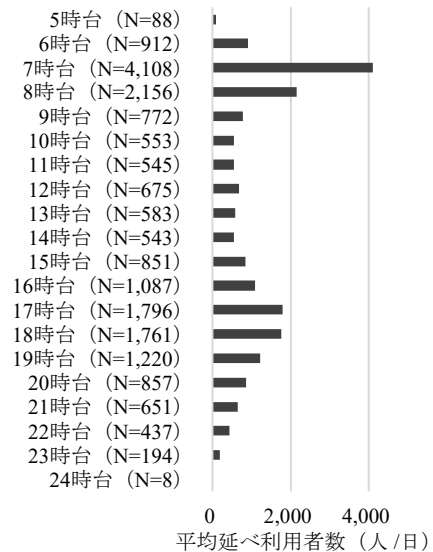


図3 時間帯別利用状況

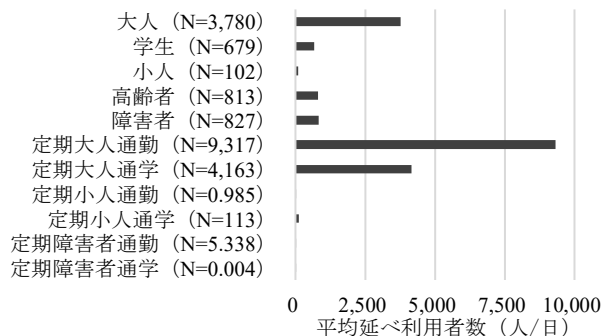


図4 利用者区別利用状況

集合のペア $G = (V, E)$ をグラフという。ただし, 無向グラフとする。

任意の 2 つの頂点 u, v について, u と v の間に辺が張られているならば, u と v は隣接しているという。 $N(u)$ を頂点 u に隣接する頂点集合とし, 頂点 u の閉近傍を $N[u] (=N(u) \cup \{u\})$ と表現する。

頂点集合 $C \subset V$ について, C の任意の 2 つの頂点が隣接しているとき, C はクリークであるという。つまり, クリークに含まれるすべての頂点間は辺で結ばれている。

3.2 移動パターン抽出のための分析手順

本研究では, どの曜日の, どの時間帯に, どのような利用者が, どの駅から, どの駅へ移動したのか, すなわち曜

日 (8 区分), 時間帯 (20 時間帯), 利用者区分 (11 区分), 乗車駅 (52 駅), 降車駅 (52 駅) の 5 次数の組合せを移動パターンと定義し, これらを同時に考慮したカード利用者の移動パターン抽出方法を考える. 本提案手法は, (1)共起グラフの構築, (2)類似グラフの構築, (3)類似グラフへのグラフ研磨の適用, (4)クリークの一挙, (5)クリーク毎に乗車駅と降車駅の組合せの把握の 5 つの処理によって構成される. 各処理について順に説明する.

3.2.1 共起グラフの構築

まず, 曜日 ($l = 1 \sim 8$) \times 時間帯 ($m = 1 \sim 20$) \times 利用者区分 ($n = 1 \sim 11$) で表現される属性頂点集合を \mathbf{X} と定義する. また, 乗車駅 ($o = 1 \sim 52$) \times 降車駅 ($d = 1 \sim 52$) で表現される OD 頂点集合を \mathbf{Y} と定義する. そして, 図 5 のように, \mathbf{X} と \mathbf{Y} の関係を表したグラフを構築する.

属性頂点 x_{lmn} には曜日, 時間帯, 利用者区分の各組合せの利用者数の情報が含まれている. 例えば, 図 5 において, 月曜日 \times 5 時台 \times 大人の属性頂点 ($x_{l=1,m=1,n=1}$) は, 月曜日の 5 時台に IruCa を利用した大人利用者の人数の情報を持つ. 一方, 各 OD 頂点 y_{od} には乗車駅, 降車駅の各組合せの利用者数の情報が含まれている. 図 5 においては, $O_1 \rightarrow D_2$ の OD 頂点 ($y_{o=1,d=2}$) は乗車駅 O_1 から降車駅 D_2 に移動した利用者数の情報を持つことになる. また, 属性頂点集合 \mathbf{X} と OD 頂点集合 \mathbf{Y} を結ぶ各辺には各属性頂点 x_{lmn} と各 OD 頂点 y_{od} の組合せでの利用者数の情報が含まれている. 例えば, 図 5 において, 月曜日 \times 5 時台 \times 大人の属性頂点 ($x_{l=1,m=1,n=1}$) と乗車駅 $O_1 \rightarrow$ 降車駅 D_2 の OD 頂点 ($y_{o=1,d=2}$) を結ぶ辺は, 月曜日の 5 時台に乗車駅 O_1 から降車駅 D_2 へ移動する大人利用者の人数を表している.

このように生成したグラフから, より特徴的な組合せのみを抽出するために, 接続関係の弱いものは削除し, 強いものだけを残す処理として, 共起グラフを構築する. ここで, 共起グラフとは, 共起性の強い頂点で構成されるグラフである. 本研究では, (1)式を用いて各属性頂点と各 OD 頂点の共起度を算出し, 共起性の強い属性頂点と OD 頂点のみを残したグラフを構築する.

$$occ(x_{lmn}, y_{od}) = \frac{|x_{lmn} \cap y_{od}|}{|x_{lmn} \cup y_{od}|} \quad (1)$$

$|x_{lmn}|$ は曜日 l , 時間帯 m , 利用者区分 n の利用者数, $|y_{od}|$ は乗車駅 o から降車駅 d の利用者数を表し, $|x_{lmn} \cup y_{od}|$ は x_{lmn} と y_{od} の和集合, $|x_{lmn} \cap y_{od}|$ は x_{lmn} と y_{od} の積集合を表す. すなわち, (1)式は各属性頂点と各 OD 頂点に含まれる利用者のうち, 共通利用者の占める割合を表現している.

次に, 属性頂点 x_{lmn} と OD 頂点 y_{od} のつながりの有無を表した 2 値変数と共起度合いを判断するための閾値 θ_c を設定し, 算出された共起度が閾値 θ_c 以上である場合は, 元のグラフの辺に含まれていた利用者数の情報を 1 に, そうでない場合は 0 に置き換える. これによって, 共起性の高い組合せのみを残した共起グラフが構成される. ここで, 新しく構築された共起グラフにおける属性頂点集合を $\mathbf{U} = \{u_i | i = 1 \sim I (I \leq 1,760)\}$, OD 頂点集合を $\mathbf{V} = \{v_j | j = 1 \sim J (J \leq 2,652)\}$ とする.

3.2.2 類似グラフの構築

類似グラフとは類似関係を表現するグラフであり, 共起グラフを元に構築する. 本研究では, 各属性頂点 $u_i \in \mathbf{U}$ の OD 頂点集合 \mathbf{V} との接続関係の類似性に着目し, 類似度を

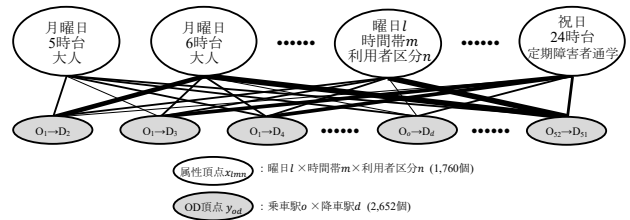


図 5 5 次数の関係を表するグラフ

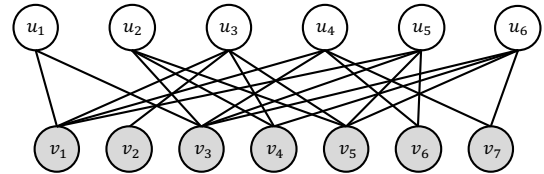


図 6-a 共起グラフ

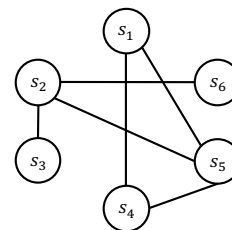


図 6-b 類似グラフ

算出することによって, 属性頂点集合による類似グラフを構築する. 類似度の算出には, (2)式を用いる.

$$sim(u_i, u'_i) = \frac{|N(u_i) \cap N(u'_i)|}{|N(u_i) \cup N(u'_i)|} \quad (2)$$

$$s. t. u_i, u'_i \in \mathbf{U}, v_j \in \mathbf{V}$$

(2)式は任意の属性頂点 u_i が隣接する OD 頂点と, 属性頂点 u'_i が隣接する OD 頂点のうち, u_i と u'_i の共通する OD 頂点占める割合を表している. 類似度はすべての属性頂点対について算出し, 閾値 θ_s を設定することによって, 類似性の強い頂点対で構成される類似グラフを構築する.

図 6-a のような共起グラフ (属性頂点集合 $\mathbf{U} = \{u_i | i = 1 \sim 6\}$ は白色の円, OD 頂点集合 $\mathbf{V} = \{v_j | j = 1 \sim 7\}$ は灰色の円) があるとすると, 属性頂点 u_1 と u_2 の類似度を算出する場合, $N(u_1) = \{v_1, v_3\}$, $N(u_2) = \{v_3, v_4, v_5\}$ であるから, $|N(u_1) \cap N(u_2)| = 1$, $|N(u_1) \cup N(u_2)| = 4$ となるので, 類似度は $sim(u_1, u_2) = 1/4 = 0.25$ となる. このようにすべての属性頂点間の類似度を算出し, 例として閾値 θ_s を 0.4 とすると, 図 6-b のような類似グラフが構成される.

類似グラフは OD 頂点の組合せが類似した属性頂点, すなわち似たような移動をする属性頂点同士を辺で結ぶことによって構成される. 例えば, 図 6-b の場合, 属性頂点 s_1 は属性頂点 s_4, s_5 と辺で結ばれているため, s_1 と s_4 は類似した移動を, s_1 と s_5 は類似した移動をとることを示している. さらに, s_4 と s_5 も結ばれていることから s_1, s_4 と s_5 の移動は類似していると考えられる. このように属性頂点間の類似関係を表現したグラフを類似グラフと呼び, 類似グラフにおける属性頂点集合を $\mathbf{S} = \{s_k | s_k \in \mathbf{U}, k = 1 \sim K (K \leq 1,760)\}$ とする.

3.2.3 類似グラフへのグラフ研磨の適用

類似グラフから, よりつながりの強い属性頂点対のみを残し, 属性頂点 $s_k \in \mathbf{U}$ をグループ化するために, 類似グラ

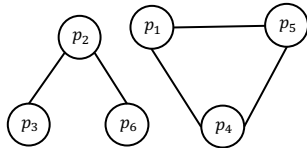


図 7 研磨グラフ

フにグラフ研磨を適用する。

グラフ研磨は、「2つの頂点に共通して隣接する頂点の個数がある程度以上多ければ、それらの頂点は一定以上の密度の部分グラフ(クリーク)に含まれている」という仮定に基づいて、グラフを構成する頂点同士の接続関係の強さから、頂点をグループ化する手法である^{9),10)}。頂点同士の接続関係の強さの評価指標には、集合の類似指標を用いる。集合の類似指標には、Simpson係数やDice係数などがあるが、本研究では既往研究と同様にJaccard係数を用いる^{9),10),12),13)}。Jaccard係数による頂点 s_k , s'_k の類似度 $sim(s_k, s'_k)$ は(3)式で求められる。(3)式は任意の頂点 s_k , s'_k の閉近傍の類似度を示している。

$$sim(s_k, s'_k) = \frac{|N[s_k] \cap N[s'_k]|}{|N[s_k] \cup N[s'_k]|} \quad (3)$$

s. t. $s_k, s'_k \in S$

グラフ研磨の計算手順について、図 6-b の類似グラフを例に説明する。まず、(3)式を用いて、類似グラフを構成するすべての属性頂点对について類似度を計算する。例えば、属性頂点 s_1 , s_2 の類似度を算出する場合、属性頂点 s_1 の閉近傍は $N[s_1] = \{s_1, s_4, s_5\}$ 、属性頂点 s_2 の閉近傍は $N[s_2] = \{s_2, s_5, s_6\}$ となるので、 $|N[s_1] \cap N[s_2]| = 1$ 、 $|N[s_1] \cup N[s_2]| = 6$ となり、類似度は $sim(s_1, s_2) = 1/6 = 0.17$ となる。

次に、閾値 θ_p を設定し、求めた類似度が閾値 θ_p 以上であれば頂点間を辺で結び、そうでなければ辺を削除する。例として閾値を0.4とすると、図 6-b グラフは図 7 のグラフのように変形される。さらに、この新たに構成されたグラフを入力グラフとして、グラフ研磨を繰り返す。この処理をグラフの変形が収束するまで行う。図 7 の例の場合、グラフ研磨を再度適用しても、グラフの形は図 7 のグラフから変化しないため、1度の適用でグラフ研磨は終了する。このようにグラフ研磨によって最終的に得られるグラフは「研磨グラフ」という。ここで、研磨グラフにおける属性頂点集合を $P = \{p_t | p_t \in S, t = 1 \sim T (T \leq 1,760)\}$ とする。

3.2.4 クリークの列挙

グラフ研磨によって得られた研磨グラフからクリークを列挙する。図 7 の研磨グラフにおいてクリークは $C_1 = \{p_2, p_3\}$, $C_2 = \{p_2, p_6\}$, $C_3 = \{p_1, p_4, p_5\}$ の3個抽出される。本提案手法によって抽出されるクリークは、似たような移動をする属性頂点 $p_t \in S$ のグループを表現し、これを考察することによって、同じ行動をする属性を把握する。

3.2.5 クリーク毎に乗車駅と降車駅の組合せの把握

抽出されたクリーク毎にどの駅からどの駅への移動が顕著であるかを把握する。例えば、月曜日×5時台×大人の属性頂点($x_{l=1, m=1, n=1}$)と月曜日×6時台×大人の属性頂点($x_{l=1, m=2, n=1}$)がクリークとして抽出された場合、最初に作成した5次数の関係を表するグラフと共起グラフを基に、属性頂点 $x_{l=1, m=1, n=1}$ および属性頂点 $x_{l=1, m=2, n=1}$ と共起性のあるOD頂点を抽出する。そして、各属性頂点とOD頂点の接続関係(乗車駅から降車駅への移動量)から、各

属性頂点に関する全利用者数のうち移動量がともに多いOD頂点を求める。これによって、何曜日の何時台のどのような利用者と、何曜日の何時台のどのような利用者の移動が類似しており、そのとき、どの駅から、どの駅へ移動したのかを把握する。

4. グラフ研磨を用いた移動パターン抽出結果

4.1 閾値の設定

本提案手法では、共起グラフ構築時に使用する閾値 θ_c 、類似グラフ構築時に使用する閾値 θ_s 、研磨グラフ構築時に使用する θ_p の3種類のパラメータが存在する。ここで、 θ_s と θ_p はともに類似度を判断するための基準なので、 $\theta_s = \theta_p$ と考えると、クリークを抽出するために必要となるパラメータは2種となる。

共起グラフ構築時の閾値 θ_c と類似グラフおよび研磨グラフ構築時の閾値 $\lambda (= \theta_s = \theta_p)$ の設定が分析結果に及ぼす傾向を把握するために、各閾値の値を少しずつ変化させ、抽出されるクリークの比較を行う。比較の際には、研磨グラフの頂点数、辺数、クリーク数、最大サイズクリークに含まれる頂点数に着目する。なお、最大サイズクリークとは、抽出されたクリークの中でクリークを構成する頂点数が最も多いクリークを指す。

共起グラフ構築時における閾値 θ_c は最大共起度0.102に対して1%、3%、5%、7%、10%を基準に5つを設定した。類似グラフおよび研磨グラフ構築時における閾値 λ については0.1から0.9まで0.1刻みで設定した。各閾値で得られた頂点数、辺数、クリーク数、最大サイズクリークの頂点数を表 1-a~1-e に示す。

まず、各閾値 θ_c (0.001, 0.003, 0.005, 0.007, 0.010)を設定したときの頂点数、辺数、クリーク数、最大サイズクリークの頂点数の違いの傾向を把握する。閾値 λ を0.1とした場合、表 1-a~1-e を比較すると、頂点数、辺数、最大サイズクリークの頂点数は閾値 θ_c を高くするにつれて、単調に減少していることがわかる。一方で、クリーク数は閾値 $\theta_c = 0.005$ で最大となり、閾値 $\theta_c = 0.001 \sim 0.005$ の間では増加傾向を、閾値 $\theta_c = 0.005 \sim 0.010$ の間では減少傾向を示している。閾値 $\lambda = 0.2$ および0.3を設定した場合でも、閾値 $\lambda = 0.1$ と同様に閾値 θ_c を高くするにつれて、頂点数、辺数、最大サイズクリークの頂点数は単調に減少し、クリーク数については最大値を持つことがわかる。しかし、閾値 $\lambda = 0.4$ および0.5を設定した場合では、閾値 θ_c を高くするにつれて、辺数と最大サイズクリークの頂点数は単調に減少しているが、頂点数とクリーク数については最大値を持つことが読み取れる。さらに、閾値 $\lambda = 0.6 \sim 0.9$ を設定した場合は、クリーク数については他の閾値 λ を設定したときと同様に、最大値を持つことが読み取れるが、頂点数、辺数、最大サイズクリークの頂点数については、減少傾向から一度増加傾向に転じていることがわかる。このように頂点数、辺数、最大サイズのクリークの頂点数の変化に一貫性がないのは、閾値 θ_c による影響だけでなく、閾値 λ も影響していると考えられる。

一方で、クリーク数については一貫した傾向を示したことから、閾値 θ_c の設定ではクリーク数の変遷に着目することとする。最大クリーク数をとる閾値 θ_c^* を中心に抽出されるクリークについて考える。まず、クリーク数の最大値をとる閾値 θ_c^* より小さい場合では、設定する閾値 θ_c が小さい

表 1-a 最大共起度に対して 1%の場合 ($\theta_c = 0.001$)

λ	頂点数	辺数	クリーク数	最大サイズ クリークの 頂点数
0.1	1,191	175,778	52	567
0.2	1,027	31,203	130	181
0.3	842	13,344	147	112
0.4	667	5,754	152	71
0.5	545	2,106	141	33
0.6	405	1,230	102	28
0.7	265	625	70	15
0.8	198	494	53	15
0.9	125	370	34	15

表 1-b 最大共起度に対して 3%の場合 ($\theta_c = 0.003$)

λ	頂点数	辺数	クリーク数	最大サイズ クリークの 頂点数
0.1	987	60,198	56	242
0.2	942	8,100	170	68
0.3	818	3,766	216	43
0.4	675	2,373	205	29
0.5	588	1,516	187	26
0.6	413	681	141	11
0.7	284	454	90	11
0.8	376	218	66	11
0.9	162	282	52	11

表 1-c 最大共起度に対して 5%の場合 ($\theta_c = 0.005$)

λ	頂点数	辺数	クリーク数	最大サイズ クリークの 頂点数
0.1	678	16,522	57	109
0.2	659	3,967	116	46
0.3	622	2,364	142	25
0.4	564	1,829	147	25
0.5	533	1,508	145	22
0.6	418	847	129	16
0.7	348	720	99	16
0.8	331	660	98	16
0.9	299	604	89	16

表 1-d 最大共起度に対して 7%の場合 ($\theta_c = 0.007$)

λ	頂点数	辺数	クリーク数	最大サイズ クリークの 頂点数
0.1	418	3,068	52	31
0.2	413	1,912	78	25
0.3	392	1,483	88	24
0.4	380	1,092	94	21
0.5	365	869	96	13
0.6	315	622	91	13
0.7	280	499	82	13
0.8	274	480	84	13
0.9	238	406	76	13

表 1-e 最大共起度に対して 10%の場合 ($\theta_c = 0.010$)

λ	頂点数	辺数	クリーク数	最大サイズ クリークの 頂点数
0.1	239	1,150	41	25
0.2	236	947	44	20
0.3	232	754	52	15
0.4	218	638	53	15
0.5	216	549	55	13
0.6	185	370	50	13
0.7	173	317	48	13
0.8	168	311	46	13
0.9	163	284	48	13

ために、共起性の低い頂点对が抽出されることになる。すなわち、本来ノイズとして削除すべき頂点对が含まれて

いるために、頂点数の多いクリークが構成され、クリークの解釈が難しくなると考えられる。一方で、閾値 θ_c^* より大きい場合では、閾値 θ_c を高くするほど、共起性の高い頂点对のみが抽出されることになる。これは決して不芳な結果であるわけではないが、ノイズに対して高感度に作用することになるため、本来抽出すべき頂点对がノイズとして削除される可能性が高い。また、閾値 θ_c 毎にすべての頂点对の共起性を比較し、ノイズの基準を設定することが考えられるが、頂点对が膨大に存在するために現実的ではなく、その判断基準は分析者に大きく依存すると考えられる。したがって、閾値 θ_c だけではノイズの判断は難しい。そこで、クリーク抽出の第一段階として、より多くのクリークを抽出することを目的に閾値 θ_c を設定することとする。すなわち、閾値 θ_c の設定ではクリーク数の最大値に着目する。

次に、各閾値 λ (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) を設定したときの頂点数、辺数、クリーク数、最大サイズクリークの頂点数の傾向を把握する。表 1-a ($\theta_c = 0.001$) において、頂点数、辺数、最大サイズのクリークの頂点数は閾値 λ を高くするにつれて、単調に減少していることが読み取れる。また、最大サイズクリークの頂点数は閾値 λ が 0.7 以上になると、15 個から変化していない。一方で、クリーク数は閾値 $\lambda = 0.4$ で最大となり、閾値 $\lambda = 0.1 \sim 0.4$ の間では増加傾向を、閾値 $\lambda = 0.4 \sim 0.9$ の間では減少傾向を示している。他の閾値 $\theta_c = 0.003 \sim 0.010$ の場合 (表 1-b ~ 1-e) を見てみると、閾値 $\theta_c = 0.003$ の場合を除いては閾値 $\theta_c = 0.001$ の場合と同様に、頂点数、辺数、最大サイズクリークの頂点数については閾値 λ を高くするにつれて単調に減少し、クリーク数については最大値を持つことがわかる。閾値 $\theta_c = 0.003$ 場合 (表 1-b) では、閾値 $\lambda = 0.8$ のときに頂点数が、閾値 $\lambda = 0.9$ のときに辺数が減少傾向から増加傾向に転じている。しかし、頂点数、辺数ともに増加数は大きくないことから、大きな変化ではないといえる。

以上の結果を踏まえて、最大クリーク数をとる閾値 λ^* を中心に抽出されるクリークについて考える。まず、閾値 λ を高くするにつれてクリークが生成される状態にあると考えられる。しかしながら、設定する閾値 λ が小さいことから、本来別々に抽出されるはずのクリークが分割されないまま、類似性の低い頂点对と類似性の高い頂点对が 1 つのクリークとして抽出されている可能性がある。一方で、クリーク数の最大値をとる閾値 λ^* より大きい場合では、どの閾値 θ_c を設定した場合でも、頂点数、辺数、クリーク数は概ね減少傾向にあったことと、最大サイズクリークの頂点数については閾値 λ を高くしても変化しなくなったことから、閾値 λ^* より大きい場合では、各クリークを構成する頂点の組合せは大きく変化しないと考えられる。すなわち、閾値 λ^* より大きい場合では、クリークの構成頂点は変化しないまま、閾値 λ を高くするにつれて生成されたクリークが削除され、類似性の高い頂点で構成されるクリークのみが抽出されると推察できる。

以上のことから、最大クリーク数をとる閾値 λ^* を設定することによって、類似グラフから適度に分割されたクリークを抽出することができると考えられる。さらに、最大サイズクリークの頂点数が変化しなくなる時の閾値 λ にも着目することで、抽出されるクリークの解釈の複雑性を防ぐことができると考えられる。したがって、本研究では最

表2-a 頂点数2個のクリーク (全65個)

	クリーク数
曜日は異なるが、時間帯・利用者区分は同じ	56
時間帯は異なるが、曜日・利用者区分は同じ	1
利用者区分は異なるが、曜日・時間帯は同じ	1
曜日・時間帯は異なるが、利用者区分は同じ	4
曜日・時間帯・利用者区分すべてが異なる	3

表2-b 頂点数3個のクリーク (全41個)

	クリーク数
曜日は異なるが、時間帯・利用者区分は同じ	31
時間帯は異なるが、曜日・利用者区分は同じ	0
利用者区分は異なるが、曜日・時間帯は同じ	0
曜日・時間帯は異なるが、利用者区分は同じ	9
曜日・時間帯・利用者区分すべてが異なる	1

表2-c 頂点数4個のクリーク (全10個)

	クリーク数
曜日は異なるが、時間帯・利用者区分は同じ	9
時間帯は異なるが、曜日・利用者区分は同じ	0
利用者区分は異なるが、曜日・時間帯は同じ	0
曜日・時間帯は異なるが、利用者区分は同じ	1
曜日・時間帯・利用者区分すべてが異なる	0

表2-d 頂点数5個のクリーク (全17個)

	クリーク数
曜日は異なるが、時間帯・利用者区分は同じ	16
時間帯は異なるが、曜日・利用者区分は同じ	0
利用者区分は異なるが、曜日・時間帯は同じ	0
曜日・時間帯は異なるが、利用者区分は同じ	1
曜日・時間帯・利用者区分すべてが異なる	0

表2-e 頂点数6個のクリーク (全2個)

	クリーク数
曜日は異なるが、時間帯・利用者区分は同じ	0
時間帯は異なるが、曜日・利用者区分は同じ	0
利用者区分は異なるが、曜日・時間帯は同じ	0
曜日・時間帯は異なるが、利用者区分は同じ	2
曜日・時間帯・利用者区分すべてが異なる	0

表2-f 頂点数9個のクリーク (全3個)

	クリーク数
曜日は異なるが、時間帯・利用者区分は同じ	0
時間帯は異なるが、曜日・利用者区分は同じ	0
利用者区分は異なるが、曜日・時間帯は同じ	0
曜日・時間帯は異なるが、利用者区分は同じ	0
曜日・時間帯・利用者区分すべてが異なる	3

表2-g 頂点数10個のクリーク (全2個)

	クリーク数
曜日は異なるが、時間帯・利用者区分は同じ	0
時間帯は異なるが、曜日・利用者区分は同じ	0
利用者区分は異なるが、曜日・時間帯は同じ	0
曜日・時間帯は異なるが、利用者区分は同じ	2
曜日・時間帯・利用者区分すべてが異なる	0

表2-h 頂点数11個のクリーク (全1個)

	クリーク数
曜日は異なるが、時間帯・利用者区分は同じ	0
時間帯は異なるが、曜日・利用者区分は同じ	0
利用者区分は異なるが、曜日・時間帯は同じ	0
曜日・時間帯は異なるが、利用者区分は同じ	0
曜日・時間帯・利用者区分すべてが異なる	1

大クリーク数をとる閾値 θ_c と、最大クリーク数をとる閾値 λ^* および最大サイズのクリークの頂点数が変化しなくなったときの閾値 λ の組合せから、各閾値を設定する方法を提案する。すなわち、本分析においては、共起グラフ構築時の閾値 θ_c は0.003、類似グラフおよび研磨グラフ構築時の閾値 λ は0.6を用いる。

4.2 属性間の類似性の把握

提案手法を交通系ICカードの利用履歴データに適用した結果、頂点数2個のクリークが65個、頂点数3個のクリークが41個、頂点数4個のクリークが10個、頂点数5個のクリークが17個、頂点数6個のクリークが2個、頂点数9個のクリークが3個、頂点数10個のクリークが2個、頂点数11個のクリークが1個の合計141個のクリークが抽出された。紙面の都合上、抽出されたクリークすべての構成内容を示すのは、困難であることから、曜日、時間帯、利用者区分の違いに着目し、得られた結果に対する考察を行う。

表2-a~2-hより、抽出されたクリークの多くは曜日や時間帯は異なるが、利用者区分は同じであることがわかった。このような組合せのクリークは、同じカード利用者の異なる曜日や異なる時間帯の移動を表現している。一方で、利用者区分のみが異なるクリークは頂点数2個のクリークで1つ抽出された。このクリークの属性頂点の組合せを見ると、「木曜日×12時台×定期小人通学」と「木曜日×12時台×定期障害者通勤」の組合せであることがわかった。この結果から、木曜日の12時台における定期小人通学利用者と定期障害者通勤利用者の移動には類似性があることがわかり、異なる利用者区分であっても同じ曜日の同じ時間帯における移動には類似性がある場合も存在することが明らかとなった。

また、曜日、時間帯、利用者区分すべてが異なるクリークは、頂点数2個のクリークにおいて3個、頂点数3個のクリークにおいて1個、頂点数9個のクリークにおいて3個、

頂点数11個のクリークにおいて1個抽出された。つまり、このような組合せのクリークは、異なる利用者間の異なる曜日、異なる時間帯間での類似性を表している。例えば、頂点数3個のクリークの組合せ、「火曜日×9時台×学生」、「水曜日×10時台×定期大人通学」、「木曜日×10時台×定期大人通学」から、定期券を持たない学生利用者の火曜日9時台における移動と定期大人通学利用者の水曜日および木曜日の10時台の移動は類似していることがわかった。また、定期券を持つ定期大人通学利用者に比べ、定期券を持たない学生利用者の移動は少し早まることが推測される。以上より、各クリークを構成する属性頂点の組合せに着目することで、類似性の高い移動をする属性が把握できることが示された。また、カード利用者数が少ない「小人」、「定期小人通学」、「定期障害者通勤」の3つの利用者区分についても、これらの利用者区分に関するクリークが、カード利用者数の多い「大人」、「定期大人通勤」、「定期大人通学」などの利用者に関するクリークと遜色なく抽

出されたことがわかった。これは、単に各属性のサンプル数の多さのみに着目するのではなく、属性間の共起性に着目した本研究の成果であるといえる。また、クリークを構成する頂点数とその組合せから、頂点数の多いクリークからは、IruCa利用者の平日の朝や夕方時間帯に行われる定常的な移動パターンが把握できることが示唆された。

4.3 IruCa 利用者の移動パターンの把握

抽出されたクリークの結果を踏まえて、クリーク毎にどの駅からどの駅への移動が特徴的であるかを把握する。これは3.2.5項で説明した手順にあたる。

抽出されたクリークは141個と数が多いため、前節と同様にすべてのクリークの乗車駅と降車駅の組合せを示すことができないため、小人（小人利用者、定期小人通学利用者）、高齢者、障害者（障害者利用者、定期障害者通勤利用者）といった交通弱者に関するクリークに着目することとする。

小人に関するクリークは15個、高齢者に関するクリークは15個、障害者に関するクリークは29個抽出された。このうち、多くは同種の利用者で構成されるクリークであったが、小人利用者と障害者利用者を含むクリークが1個、定期小人通学利用者と定期障害者通勤利用者を含むクリークが3個あることがわかった。以降、これら4個の異なる利用者区分で構成されるクリークに注目し、最も移動量の多い乗車駅と降車駅の組合せを把握する。

着目する4個のクリークの構成内容は、

(C1) 「祝日×8時台×小人、土曜日×23時台×障害者」、

(C2) 「木曜日×12時台×定期小人通学、木曜日×12時台×定期障害者通勤」、

(C3) 「火曜日×8時台×定期小人通学、水曜日×8時台×定期小人通学、木曜日×8時台×定期小人通学、金曜日×8時台×定期小人通学、金曜日×9時台×定期小人通学、月曜日×13時台×定期障害者通勤、火曜日×13時台×定期障害者通勤、水曜日×13時台×定期障害者通勤、金曜日×13時台×定期障害者通勤」、

(C4) 「火曜日×9時台×定期障害者通勤、金曜日×10時台×定期障害者通勤、月曜日×14時台×定期小人通学、月曜日×14時台×定期障害者通勤、火曜日×14時台×定期小人通学、火曜日×14時台×定期障害者通勤、水曜日×14時台×定期小人通学、水曜日×14時台×定期障害者通勤、木曜日×14時台×定期障害者通勤、金曜日×14時台×定期小人通学、金曜日×14時台×定期障害者通勤」

である。この(C1)から(C4)までの4個のクリーク毎に、各クリークを構成する属性頂点とOD頂点の接続関係から各属性頂点に関する利用者のうち最も利用者数の多いOD頂点を求めることによって、各クリークの特徴的な乗車駅と降車駅の組合せを抽出する。各クリークにおいて移動量の多い乗車駅と降車駅の組合せを表3に示す。

表3より、抽出された乗降車駅はすべて長尾線の駅であることがわかった。(C1)の乗車駅と降車駅の組合せの結果から、祝日の8時台の小人利用者と土曜日の23時台の障害者利用者には類似性があり、林道駅から公文明駅へ移動する傾向にあるといえる。

(C2)の結果からは木曜日の12時台の定期小人通学利用者と定期障害者通勤利用者は木太東口駅から白山駅への移動が顕著であることがわかった。

表3 移動量の多い乗車駅と降車駅の組合せ

クリーク	乗車駅	降車駅
(C1) 祝日×8時台×小人, 土曜日×23時台×障害者	林道	公文明
(C2) 木曜日×12時台×定期小人通学, 木曜日×12時台×定期障害者通勤	木太東口	白山
(C3) 火曜日×8時台×定期小人通学, 水曜日×8時台×定期小人通学, 木曜日×8時台×定期小人通学, 金曜日×8時台×定期小人通学, 金曜日×9時台×定期小人通学, 月曜日×13時台×定期障害者通勤, 火曜日×13時台×定期障害者通勤, 水曜日×13時台×定期障害者通勤, 金曜日×13時台×定期障害者通勤	木太東口	長尾
(C4) 火曜日×9時台×定期障害者通勤, 金曜日×10時台×定期障害者通勤, 月曜日×14時台×定期小人通学, 月曜日×14時台×定期障害者通勤, 火曜日×14時台×定期小人通学, 火曜日×14時台×定期障害者通勤, 水曜日×14時台×定期小人通学, 水曜日×14時台×定期障害者通勤, 木曜日×14時台×定期障害者通勤, 金曜日×14時台×定期小人通学, 金曜日×14時台×定期障害者通勤	長尾	木太東口

(C3)と(C4)の乗車駅と降車駅の組合せを見てみると、ともに木太東口駅と長尾駅の組合せである。また、2つのクリークは「定期小人通学」と「定期障害者通勤」に関するクリークであり、曜日と利用者区分については組合せが類似しているといえる。このことから、平日8~9時台の定期小人通学利用者および13時台の定期障害者通勤利用者は木太東口駅から長尾駅へ移動し、14時台になると長尾駅から木太東口駅へ帰宅行動をとると考えられる。

以上の結果から、クリーク毎に移動量の最も多いOD頂点を抽出することによって、曜日、時間帯、利用者区分間で類似性が高い組合せがわかり、それら利用属性をもつ利用者が、どの駅から、どの駅へ移動しているのかを把握することが可能となった。

5. おわりに

本研究では、グラフ研磨を用いた交通系ICカードデータからのパターン抽出の方法論について提案した。具体的には、(1)共起グラフの構築、(2)類似グラフの構築、(3)類似グラフへのグラフ研磨の適用、(4)クリークの列挙、(5)クリーク毎に乗車駅と降車駅の組合せの把握の5つの処理で構成される方法論を提案し、グラフ研磨の交通系ICカードデータ分析への適用可能性を検討した。

実証分析に際して、香川県で運行されている鉄道路線「ことでん」で利用可能な交通系ICカード「IruCa」の利用履歴データを用いた。有効データ9,008,709件から、曜日(8区分)、時間帯(20時間帯)、利用者区分(11区分)、乗車駅(52駅)、降車駅(52駅)の関係を表すグラフを構築し、このグラフに対して本提案手法を適用した。以下に、本研究の成果をまとめる。

本研究では、類似グラフへのグラフ研磨の適用によって、より接続関係の強い属性頂点をグループ化した。さらに、抽出されたクリークから属性間の類似性を把握し、類似した移動をとると考えられる属性グループを明らかにした。抽出されたクリーク 141 個の結果から、多くのクリークは同一利用者区分で表現されるクリークであることがわかった。一方で、曜日、時間帯、利用者区分のすべてが異なる組合せのクリークからは、異なる利用者区分間に類似性があることを定量的に明らかにすることができることが示唆された。IruCa のような地域限定型の交通系 IC カードデータにはカード利用者の日常的な移動情報が大量に含まれているために、同一利用者区分で表現されるクリークが多くなることは十分に想定できる。しかし、異なる曜日、異なる時間帯、異なる利用者区分で表現される属性頂点の組合せを事前に予測することは難しい。例えば、本提案手法によって「祝日×8 時台×小人」と「土曜日×23 時台×障害者」の属性頂点によって構成されるクリークが抽出されたが、このような組合せは、集計分析のような基礎的な分析では把握できない。各利用者区分の利用者毎に類似関係のある曜日と時間帯の組合せを見つける方法も考えられるが、考慮しなければならない回数が増えるほど、その組合せ数は莫大になり、現実的な抽出手法とはならないと考えられる。したがって、「祝日×8 時台×小人、土曜日×23 時台×障害者」のような組合せは、本提案手法のように複数の回数を同時に考慮できる分析手法を用いなければ、わかり得なかった結果であるといえる。さらに、クリーク毎に抽出した乗車駅と降車駅の組合せの結果から、交通系 IC カード利用者の主特徴挙動として、どの曜日の、どの時間帯に、どのような利用者が、どの駅から、どの駅へ移動しているのかを把握することができることがわかった。

以上の結果から、本提案手法を用いることによって、8 曜日区分×20 時間帯×11 利用者区分×52 乗車駅×52 降車駅の組合せを持つグラフから、IruCa 利用者の移動パターンを抽出できることが示された。また、カード保有者数の少ない小人、高齢者、障害者といった交通弱者の利用を表すクリークは、抽出された全クリークの約 40% を占め、本提案手法を用いることによってサンプル数に依存することなくパターン抽出が可能であることが示唆された。

さらに、サンプル数や考慮する変数の組合せ数などが大きいデータであっても、共起グラフ構築時の閾値 θ_c と類似グラフおよび研磨グラフ構築時の閾値 λ を設定することによって、分析結果の複雑性を防ぐことが可能であるとわかった。また、変数間の共起性や類似性を考慮していることから、サンプル数が少ないために抽出されにくいデータの潜在的特徴も明らかにできることがわかった。

本提案手法では交通系 IC カードデータからの移動パターン抽出のための手順として、5 つの処理を設定したが、計算手順の複雑さをなくし、より少ない処理工程で、効率的なアルゴリズムを考える必要がある。また、他の最新の分析手法と本提案手法の分析結果を比較し、結果の解釈の複雑性や抽出されるパターンの傾向の違いを検証することによって、本提案手法がパターン抽出方法としてどの程度有効性があるのかを検討する必要がある。

謝辞

本研究の遂行にあたり、高松琴平電気鉄道株式会社様か

ら、大変貴重なデータをご提供いただいた。ここに、感謝の意を表す。香川高等専門学校の宮崎耕輔准教授、東京大学生産技術研究所の伊藤昌毅講師には、データの収集および処理に際して、ご尽力いただいた。記して謝意を表す。

参考文献

- [1] 服部 誠一郎, 小塚 みすず, “交通弱者 (年少者) の潜在的な交通需要に関する研究”, 日本都市計画学会関西支部研究発表会公演概要集, Vol.14, pp.93-96 (2016).
- [2] 高木 昌也, 奥嶋 政嗣, 近藤 光男, “地方都市のバス利用促進のための行動意図法による潜在需要分析”, 都市計画論文集, Vol.45-3, pp.439-444 (2010).
- [3] 永井 政伸, 日比野 直彦, 森地 茂, “ETC-OD データを用いた都市高速道路における経路選択行動に関する研究”, 土木学会論文集 D3, Vol.67, No.5, pp.1_589-I_597 (2011).
- [4] 田中 伸治, 藤原 直生, 桑原 雅夫, “携帯電話からの GPS データを用いた歩行者の経路選択リンク集合の推定”, 土木学会論文集 D3, Vol.67, No.5, pp.1_617-I_624 (2011).
- [5] 樋口 彰, 服部 宏充, “プローブカーデータに基づいた京都市観光者の観光行動分析”, 人工知能学会全国大会論文集, Vol.28, CD-ROM (2014).
- [6] 嶋本 寛, 北脇 徹, 宇野 伸宏, 中村 俊之, “IC カード利用履歴データを用いた公共交通需要変動分析”, 土木学会論文集 D3, Vol.70, No.5, pp.1_605-I_610 (2014).
- [7] 野上 祐人, 片岡 源宗, 熊谷 靖彦, “IC カード「ですか」を活用した高知中央地域の公共交通利用実態の基礎分析”, 土木計画学研究・講演集, Vol.44, CD-ROM (2011).
- [8] 日下部 貴彦, 朝倉 康夫, “交通系 IC カードデータによる鉄道利用者行動のバスケット分析”, 土木計画学研究・講演集, Vol.41, CD-ROM (2010).
- [9] 宇野 毅明, 中原 孝信, 前川 浩基, 羽室 行信, “データ研磨によるクリーク列挙クラスタリング”, 情報処理学会アルゴリズム研究会報告書, Vol.146, No.2, pp.1-8 (2014).
- [10] 中原 孝信, 羽室 行信, 宇野 毅明, “グラフ研磨手法を用いた顧客の店舗選択モデル”, オペレーションズ・リサーチ, Vol.60, No.2, pp.89-95 (2015).
- [11] ことでんグループ, ことでん路線図, <http://www.kotoden.co.jp/publichtm/kotoden/time/index.htm>.
- [12] 中原 孝信, “グラフ研磨による考慮集合の予測と推薦システム”, http://www.orjs.or.jp/archive2/or61-11/or61_11_762.pdf (2016).
- [13] 前川 浩基, 内田 将史, 大内 章子, 宇野 毅明, “羽室 行信, データ研磨手法を用いた Twitter ユーザの関係構造変化の検出”, 人工知能学会全国大会論文集, Vol.28, CD-ROM (2014).