

スポット重要度に基づく適応的变化検出

尾ノ上 晃^{1,a)} 堀 磨伊也¹ 島田 敬士¹ 谷口 倫一郎¹

概要：本研究では、グラフ理論における中心性を用いて求めたスポットの重要度に基づく、人流傾向の変化の適応的な検出手法を提案する。中心性は新たに提案する SpotRank を用いることにより、各スポット間の移動人数と流出入関係も考慮に入れたスポットの重要度を推定する。これにより、各スポット間の移動人数のみからでは求めることが困難な実際の需要に合った解析エリアにおける重要スポットの検出が可能となる。また、変化検出は、変化の度合いを知りたい目的の事象と比較対象となる事象を、ユーザがそれぞれ柔軟に作成可能なシステムを作成する。比較対象となるデータの分布はノンパラメトリックなモデルによる推定法であるカーネル密度推定を用いることで、各スポットでの異なる人流傾向の分布に対応する。

Adaptive Change Detection Based on SpotRank

Akira Onoue^{1,a)} Maiya Hori¹ Atsushi Shimada¹ Rin-ichiro Taniguchi¹

1. はじめに

本研究では、グラフ理論の中心性を用いて求めたスポットの重要度に基づく、人流傾向の変化の適応的に検出する手法を提案する。解析の対象となるエリアや施設において、人々の行動がどう変化しているかという情報は、様々な用途に活用可能である。例として、各観光地の関係や観光客の動向の把握、広告配信やイベント開催などのマーケティング、災害時の避難誘導などの際に有益な情報となることが期待される。人流に関する従来の変化検出は、単一スポットに存在した人数の変化や、スポット間の移動人数の変化を検出するもの [1] が存在する。しかし、多数のスポットとスポット間を移動する人々の構造はネットワークと捉えることが可能である。よって、単一スポットのみに注目するのではなくネットワーク全体の中での一スポットと考え分析し、変化を検出すべきである。

人流解析のシステムは、解析対象であるエリアや施設の管理者等のユーザが目的に応じてその結果を容易に取得、理解可能であることが望ましい。そのため本研究では、変化の度合いを知りたい目的のデータと比較対象となるデータを、ユーザがそれぞれ柔軟に作成可能である変化検出を行う。これを適応的变化検出と呼ぶこととする。比較対象

となるデータの分布はノンパラメトリックなモデルによる推定法であるカーネル密度推定を用いることで、各スポットでの異なる傾向の分布に対応する。

また、中心性は、インターネット上の Web サイト同士の関係や SNS ユーザ同士の関係の分析 [2][3] のみでなく、実世界における交通網の分析 [4][5] などにも用いられている。本研究では、多地点間における人の移動を表す人流ネットワークに中心性指標を導入し、各スポットの重要度を推定する。スポットの重要度の推定は L. Page ら [6] の PageRank の手法を基に作成し、これを SpotRank と呼ぶこととする。SpotRank を用いることにより、各スポット間の移動人数と流出入関係も考慮に入れたスポットの重要度を推定し、各スポット間の移動人数のみからでは求めることが困難な、周辺スポットとの関係性を考慮した解析エリアにおける重要スポットの検出が可能である。

なお、本研究におけるスポット間の移動人数は、Wi-Fi パケットセンサを用いて収集された Probe Request 情報から得られたユーザ数とする。Probe Request 情報とはスマートフォン等の端末が発信する Wi-Fi アクセスポイントへの接続要求である。Probe Request 情報はユーザが端末に特定のアプリをインストールする必要も無い。また、端末の Wi-Fi 機能をオンにした状態であれば収集可能である。このように、他の手法に比べ、広範囲における不特定多数の人々の行動を分析するのに適している。

¹ 九州大学

^{a)} onoue@limu.ait.kyushu-u.ac.jp

2. 関連研究

ネットワーク分析により、ネットワーク全体の特徴や複数ノード間の関係を把握することが可能である。Web上のハイパーリンク構造や企業間の関係 [7] だけでなく実世界の環境に適用させた分析も行われている。実世界におけるスポットの分析を行う際、単一スポットのみに注目して分析を行うのではなく、周辺スポットとの関係を含めた分析を行うことで、単一スポットのみに注目した分析では得ることができない知見が得られると考えられる。複数スポットと道路ネットワークや人々の移動ネットワークに関する研究も行われている。以下、それらを対象としたスポットの重要度に関する研究について説明する。

2.1 構造のみに基づくネットワーク分析

中心性を用いて都市や公園の道路ネットワークにおいて各スポットの重要度を評価する研究が存在する。観光地における最適な施設配置場所の把握やテーマルートの計画、インフラ整備などの際に、対象エリアの構造を定量的に評価することが可能となる。H. Shih[8]はドライブ観光の観光地ネットワーク分析を行っている。各スポット間の移動台数がしきい値より多い場合にスポット間に枝を張ることで、ネットワークを構成する。次数中心性、近接中心性、媒介中心性の3つの指標で分析を行っている。その結果、重要度の高いスポットには観光案内所等の施設が設置されている傾向にあり、ネットワーク分析の手法が実環境における人々の行動に基づくスポットの分析に応用可能であることを示している。また、K. Taczanowskaら [9]は、中心性を用いて広域の公園におけるハイキングコースの構造と利用状況の評価する手法を提案している。物理的ネットワーク（無向グラフ）と利用者の移動方向も含めた機能ネットワーク（有向グラフ）について、それぞれ次数中心性、近接中心性、媒介中心性の3つの指標で分析を行っている。その結果、施設管理上の観点からは機能ネットワークで得られた結果が特に重要であることを示している。この手法により、従来のGISベースのネットワーク解析と組み合わせることで補間可能であるとしている。

しかし、構造のみに基づくネットワーク分析によって各スポットの重要度を求める場合、実世界特有の各スポットの特性や各スポット間の距離等の問題を考慮していない。

2.2 実世界特有の性質を考慮したネットワーク解析

各スポットの特性や各スポット間の距離等の実世界特有の性質を考慮したネットワーク分析に関する研究も行われている。T. Agrtzkovら [10]は、中心性を求める手法の一つであるPageRank[6]を用いて、都市の道路ネットワークにおける各交差点の重要度を求めている。この際、外的要因をノード（交差点）の特性とし、交差点周辺に店舗等の



図 1: スポット間移動の可視化（暖色なほど人数が多い）

施設が存在する場合、それらを定量化し考慮することで重要度を算出する手法を提案している。従来のPageRankとの比較の結果、都市の道路ネットワークの各交差点の重要度を算出する際の、外部要因を考慮する重要性が示されたとしている。伏見ら [11]は、重要観光スポットの抽出を目的とし、実距離を考慮した中心性指標の提案を行っている。距離を考慮することにより、目的に沿ったスポットの抽出が行えることを示している。

これらのことから実世界特有の性質を考慮したネットワーク解析が有用であると示唆されている。しかし、スポットの特性やスポット間の距離がスポットの重要度に影響する度合いには個人差がある。また、各スポットを実際に訪れるのは人々であるため、施設の存在や距離を考慮するのではなく、各スポット間の移動人数を考慮することにより、より実際の事象にあった結果が得られると考える。

3. 提案手法

様々なセンサを用いて、人々の行動分析は盛んに行われている。また、図 (1) のように、移動人数や傾向の可視化も可能となった。本研究では、人々のスポット間移動推定を、Wi-Fi パケットセンサを用いて行う。その手法は、尾ノ上ら [12]の手法を用いる。以下、3.1 節で提案手法の基となるPageRank アルゴリズムについて、3.2 節で提案手法であるSpotRank アルゴリズムについて述べる。最後に、3.3 節で適応的变化検出について述べる。

3.1 PageRank

PageRankとは、L. Pageら [6]によって提案された、インターネット上のWebページのハイパーリンク構造から各Webページの重要度を算出するアルゴリズムである。PageRankは大きく分けて以下の3つで決定される。

- 被リンク数
- 重要なページからのリンクであるか
- リンク元ページのリンク数

つまり、より多くのリンクを受けるページほど重要である。重要なページからのリンクを受けるページもまた重要である。さらに、リンク数の少ないページからのリンクは貴重なリンクであるとし、重要度は高くなる。以下に C. Meyer ら [13] による PageRank のアルゴリズムを説明する。

Web ページのハイパーリンク構造は有向グラフである。以下、 n 個のページによるハイパーリンク構造を考える。各ページをノードとし、ページ p_i から p_j にリンクが存在する場合、 p_i から p_j エッジを張る。 c_i をページ i からのアウトリンク数（ページ i から他ページへのリンクの総数）とする。 Web のハイパーリンク構造を $n \times n$ 正方行列 \mathbf{H} で表す。この時、行列 \mathbf{H} の成分 a_{ij} は式 (1) で定義される。

$$h_{ij} = \begin{cases} \frac{1}{c_i} & (c_i \neq 0) \\ 0 & (c_i = 0) \end{cases} \quad (1)$$

つまり、アウトリンク数が 0 のノード（以下、ダングリグノード）の行の成分は全て 0、非ダングリグノードの行の成分の和は 1（行方向に正規化）となる。ここで各ページの重要度を成分とする $1 \times n$ の PageRank ベクトルを \mathbf{r}^T とすると k 回目の反復計算時の PageRank ベクトルは、

$$\mathbf{r}^{(k+1)T} = \mathbf{r}^{(k)T} \mathbf{H} \quad (2)$$

と表される。式 (2) を $\mathbf{r}^{(k+1)T}$ が \mathbf{r}^T 収束するまで反復計算を行い、解を求める。しかし、ダングリグノードが存在する場合、行列 \mathbf{H} のみを用いた反復計算では \mathbf{r}^T が収束しない可能性がある。行列がマルコフ連鎖の確率行列であるとき、全ての成分が非負であり、各行の合計は 1 であるので、行列 \mathbf{H} が確率行列となるためにこれを修正し、ハイパーリンク行列 \mathbf{S} とする。

$$\mathbf{S} = \mathbf{H} + \frac{\mathbf{a}e^T}{n} \quad (3)$$

ここで e は全ての成分が 1 の列ベクトル、 \mathbf{a} は式 (4) で表される列ベクトルである。

$$\mathbf{a}_i = \begin{cases} 1 & (\text{ページ } p_i \text{ がダングリグノード}) \\ 0 & (\text{ページ } p_i \text{ が非ダングリグノード}) \end{cases} \quad (4)$$

さらに、 \mathbf{r} が反復計算によりただ一つの定常ベクトルに収束することを保証するために、ハイパーリンク行列 \mathbf{S} を修正し、正かつ既約な確率行列であるグーグル行列 \mathbf{G} を定義する。

$$\mathbf{G} = d\mathbf{S} + (1-d)\mathbf{E} \quad (5)$$

ここで、 $\mathbf{E} = \mathbf{e}e^T/n$ 、また、 $d(0 \leq d \leq 1)$ はダンピング・ファクターと呼ばれる、ノードに存在するエッジにより遷移する場合とランダムに遷移する場合の割合を表すパラメータであり、通常 $d = 0.85$ に設定される [13]。よって、

グーグル行列 \mathbf{G} を用いて式 (2) を再定義すると、

$$\mathbf{r}^{(k+1)T} = \mathbf{r}^{(k)T} \mathbf{G} \quad (6)$$

となり、反復計算により収束し、最大固有値 1 に対応する固有ベクトル（PageRank ベクトル） \mathbf{r}^T を求めることが可能である。

3.2 移動人数を考慮した中心性指標

各スポット間の移動人数も考慮した中心性指標である SpotRank を提案する。本研究では、ネットワークにおける各 Web ページの重要度を評価する PageRank アルゴリズム [6][13] における Web ページを、実世界のスポットに置き換えることで各スポットの重要度を求める。対象のスポットを含む解析範囲において、スポットをノード、時刻間 t で移動した人が存在するスポット間にエッジを張ることでネットワークの構造を決定する。ここでは、[10] や [11] で行われているような道路構造に基づくネットワーク構造の決定は行わない。そのため、同一の解析範囲である場合でも、人々の移動状況によって、その構造が変化する動的なネットワークであることに注意されたい。このネットワークにおけるスポットから他のスポットへ移動することを流出、他のスポットから移動してくることを流入とする。このとき、SpotRank は以下の 3 要素で決定される。

- スポットへの流入経路数
- 重要なスポットからの流入であるか
- 流入元スポットの流出経路数

基本的な考え方は、多くのスポットから人が流入しているスポットの重要度は高いとする。また、重要度の高いスポットからの流入が存在するスポットもまた重要度が高いとする。さらに、流出経路数が多いスポットからの人の流入経路より、数少ない流出経路を持つスポットからの流入経路の方が貴重な経路であるとして重要度が高くなる。加えて、本研究では、流入人数が多い経路ほど重要な経路という考え方に基づき、スポットからの流出経路ごとに流出人数に応じて重み付けを行う。以下、各スポットの重要度を成分とする $1 \times n$ の SpotRank ベクトル \mathbf{s}^T を求める手法を説明する。

Wi-Fi パケットセンサにより収集されたデータから、OD 表（Origin Destination Table）と呼ばれる $n \times n$ 行列 \mathbf{M} を作成する。OD 表は行が出発スポット、列が到着スポットであり、 \mathbf{M} の (i, j) 成分 m_{ij} は時刻間 t にスポット i からスポット j へ移動人数を表す。

$$M'_i = \frac{M_i}{\sum_{k=1}^n M_{ik}} \quad (7)$$

式 (7) に示すように、行列 \mathbf{M} の各行 M_i を各行の合計人数で割った行列を \mathbf{M}' とする。流出人数が 0 人のスポット（PageRank におけるダングリグノード）の行の成分は全

て0, 流出する人が存在するスポットに対応する行の成分の和は1である. 行列 M' が, PageRankにおける行列 H と異なる点は, 行列 H は, 式 (1) で示されるように, あるページからの全アウトリンクに対して等しい遷移確率を与えているのに対し, 行列 M' は, 流出人数が多い経路ほど高い遷移確率を与えている点である. これにより, 移動人数を考慮した各スポットの重要度の算出が可能となる. 次に, SpotRank のハイパーリンク行列を S' とし式 (8) で定義される.

$$S' = M' + \frac{ae^T}{n} \quad (8)$$

ここで, a , e は式 (3) と式 (4) で与えられたベクトルである. 次に, 行列 S' を用いて新たな行列 X を式 (9) 定義する.

$$X = dS' + (1-d)E \quad (9)$$

行列 X の固有値1に対応する固有ベクトルを, 式 (10) を反復計算することにより求める.

$$s^{(k+1)T} = s^{(k)T} X \quad (10)$$

この固有ベクトルが, SpotRank ベクトル s^T である.

3.3 通常モデルの適応的作成による変化検出

人流傾向の変化の適応的な検出手法として, 変化の度合いを知りたい目的の事象と比較対象となる事象を, ユーザがそれぞれ柔軟に選択可能なシステムを作成する. ユーザが各事象を柔軟に選択可能なシステムである理由としては, 同一目的の事象の場合であっても, 比較対象となる事象を選択する際に用いるデータの選択条件により結果は大きく変化する場合がある点が挙げられる. 人流解析の分野においても, ある期間の事象に対して様々な事象との比較によりその期間の人流はどのような特徴を持つのかを推定可能である. 事象の選択条件を固定するのではなく, 柔軟に選択可能にすることで, ユーザの目的によって何をもって変化とするかが異なる場合に対応する.

変化の度合いを知りたい目的の事象を表すデータを, 時刻間 t を1フレームとする目的データ D^t , 比較対象である通常状態とする事象のデータを, 同一幅 t の n フレームのデータの集合として通常モデル D^b とする. 通常モデルを構成する s 個目のデータを通常データ $D^{b(s)}$ ($1 \leq s \leq n$) とする. 通常モデルを式 (11) に示す.

$$D^b = \{D^{b(1)}, D^{b(2)}, \dots, D^{b(s)}, \dots, D^{b(n)}\} \quad (11)$$

また, 各データの関係性を図2に示す.

通常データの選択条件によって, 通常モデルの分布も異なる. そのため, 本研究では, 有限個の標本から, その標本を生成する確率密度関数を推定するノンパラメトリック法であるカーネル密度推定法 [14] を用いる. 標本を $\{x_i\}_{i=1}^n$ とすると, 確率密度関数 $f(x)$ の推定量は,

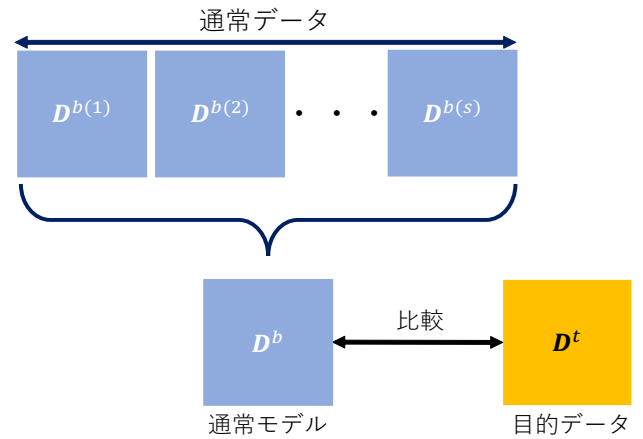


図2: 通常モデルを用いた変化検出手法

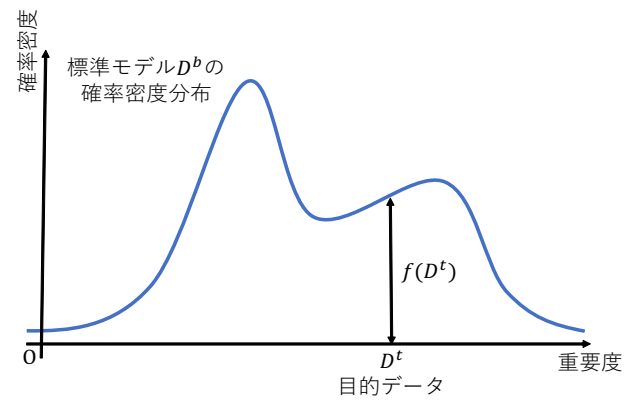


図3: 変化の度合いの定義

$$\hat{f}_{KDE}(x) = \frac{1}{n} \sum_{i=1}^n K(x) \quad (12)$$

で表される. 式 (12) における $K(x)$ はカーネル関数であり, 本研究では, 1次元の標準ガウスカーネル関数

$$K_{\sigma}(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \quad (13)$$

を用いる. 式 (13) において, σ はカーネルのバンド幅である. バンド幅はハイパーパラメータであり, カーネル密度推定法の推定精度はカーネル幅に依存する. 本研究では, 交差確認法によって最適なカーネル幅を選択する.

目的データの通常モデルからの変化の度合い $y(D^t)$ は, カーネル密度推定によって求めた確率密度分布における目的データの情報量

$$y(D^t) = -\ln f(D^t) \quad (14)$$

によって定義する (図3). 情報理論における情報量は, 目的データの事象の不確定さの減少量を表す. よって, 発生する確率が低いほど情報量が多いという考え方は, 変化の度合いとして定義するのに適している.



図 4: Wi-Fi パケットセンサ設置場所

表 1: Wi-Fi パケットセンサを 2 台設置しているスポット

スポット	階	施設名	用途
R_1	1	メインダイニング	食堂
C_1	1	クアシス	カフェ
L_1	4	鷗鳴天空広場	自習スペース
R_2	B1	ビッグダイニング	食堂

4. 実験

提案手法の有用性を示すため、SpotRank は、同一データによって作成されたネットワークに PageRank[13] を適用した結果との比較を行う。適応的変化検出については、通常モデルの選択条件により検出結果の違いを比較し、結果に違いがみられるかを検証する。また、その原因について考察する。

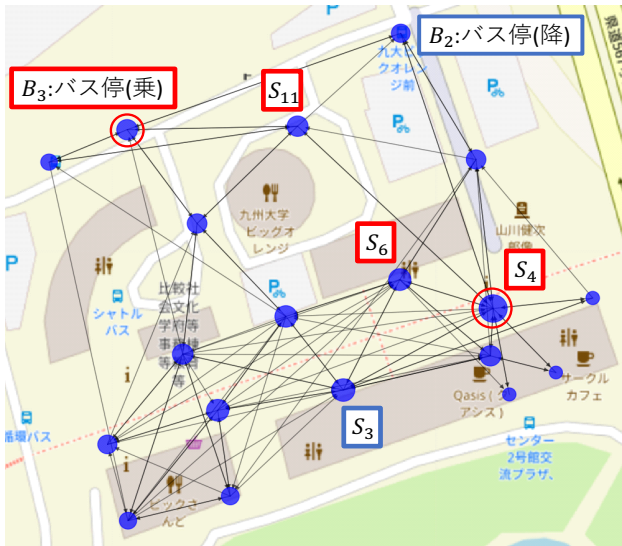
4.1 Wi-Fi パケットセンサでのデータ収集期間、解析範囲

分析対象となるデータは 2016 年 5 月 10 日から 2017 年 9 月 17 日の期間に、九州大学伊都キャンパス内の 20 地点 (以下、スポット) に設置された計 24 台の Wi-Fi パケットセンサで収集された Probe request 情報である。分析対象となった端末数は 165,893 台であった。図 4 に Wi-Fi パケットセンサの設置場所 (スポット) を示す。各スポットの記号はスポットの特性を表す。R は食堂、C はカフェ、B はバス停、L は教室や自習スペース、S は無特性なスポットである。スポット L_2 , L_3 は平成 29 年 4 月 1 日に Wi-Fi パケットセンサが設置されたため、それ以降のデータを解析対象とした。また、センター 1 号館とセンター 2 号館は講義が行われる棟である。1 スポットにつき Wi-Fi パケットセンサが 2 台ずつ設置されている箇所については表 1 にまとめた。この 4 スポットにおいては、2 台の Wi-Fi パケットセンサをまとめて同一のスポットとして扱った。

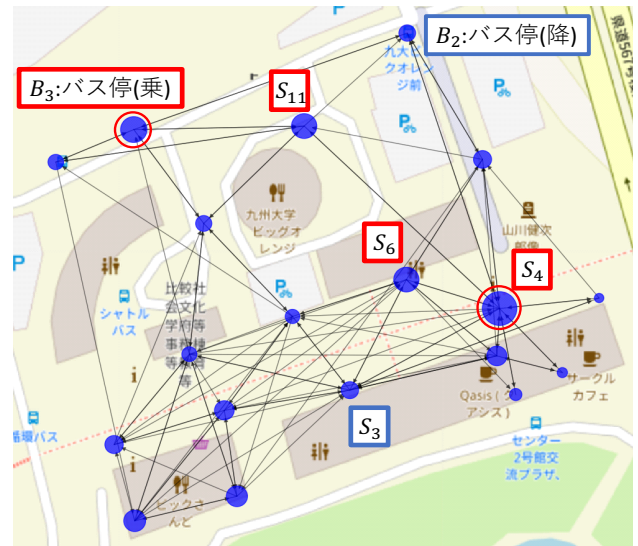
4.2 SpotRank の評価

提案手法である SpotRank と、PageRank[13] によるスポットの重要度の結果を比較する。これにより、SpotRank を用いることで、指定した時間帯において中心的役割を果たしていると考えられるスポットに対して、高い重要度が与えられているかを検証した。なお、解析範囲が大学キャンパスであるという特性上、平日と休日で異なる人流の傾向となる。また、新歓行事や学祭などのイベントもその傾向に大きく影響する。そのため、SpotRank の結果と、実際の事象との比較を行いやすくするために、特別なイベントも行われず、大学において通常の授業日と考えられる平日のデータを用いた評価を行う。本研究では、2017 年 5 月 2 日のデータを用いた。

図 5 に、2017 年 5 月 2 日の 13 時から 1 時間のデータを用いた場合の、ネットワークを地図上に示す。ノードの大きさがスポットの重要度を表し、重要度が高いほど、ノードは大きくなる。また、エッジは人の移動が存在したスポット間のみに張った。図 5 (a) に PageRank を用いた結果を示す。PageRank を用いた場合、重要度の高いスポットは、スポット S_3 , S_4 , S_6 などであり、他の多くのスポットからの流入経路を持っていることが分かる。これは、スポット S_3 , S_4 , S_6 は講義棟の入り口、またはその近くに位置しており、13 時は昼休み終了後、午後の講義を受けるため、他の多くのスポットから人々が流入する。そのことから、実際の現象に適した結果であるといえる。図 5 (b) に SpotRank を用いた結果を示す。SpotRank を用いた場合、重要度の高いスポットは、スポット B_2 , S_4 , S_{11} などであった。PageRank と同じ、スポット S_4 の重要度が高い。これは、スポット S_4 が多くの流入経路を持っているのに加えて、各経路からの流入人数も多かったことが原因と考えられる。また、スポット S_4 は、食堂やカフェ、バス停降り場から講義棟方面へ向かう人々の多くが通る。よって、SpotRank においても重要度が高いのは妥当であり、当該時間帯におけるハブスポットであるといえる。PageRank と SpotRank の結果で異なる点としては、スポット S_{11} , B_2 の重要度が高い点である。スポット B_2 は大学から駅方面へ向かうバスの乗り場である。13 時台は、午前中の講義を受けた学生が、帰宅するためにバスを利用する時間帯であることから、当該時間帯において重要なスポットであるといえる。キャンパスの構造上、食堂や講義棟からバス停に向かう経路は限られているため、同一の経路を多くの人が利用した。そのため、バス停は流入経路数が少ないが流入人数は多いという特徴を持ち、PageRank では、重要度が低くなったのに対し、SpotRank では、重要度が高くなった。また、スポット S_{11} はバス停へ向かう経路上に存在するため、重要度が高くなったと考えられる。よって、SpotRank を用いた結果は、実際の施設の利用状況に適しているといえる。



(a) PageRank



(b) SpotRank

図 5: ネットワーク構造と重要度 (2017/05/02 (金) 13 時)

4.3 適応的变化検出の結果

次に、SpotRank によるスポットの重要度を用いた適応的变化検出の結果を示す。解析対象の日は、2017 年 1 月 11 日とした。以下にその理由を述べる。本キャンパスにおいて、同一データを用いて、日単位での変化検出が行われている [12]。水曜日を対象に通常モデルとの変化の度合いを求めた結果、複数日において変化の度合いが大きい日が検出された。その多くの日において、イベント開催等の変化の度合いが大きくなった原因が考えられたのに対し、2017 年 1 月 11 日はイベント開催等のない通常の平日であったにもかかわらず、変化の度合いが大きかった。よって、解析対象の日を 2017 年 1 月 11 日とし、提案手法を用いて、より詳細な変化検出を行う。なお、本研究は、通常モデルを適応的に選択することで単一の通常モデルのみを用いた変化検出では、得ることができない知見を得ることを目的とする点で [12] と異なる。本研究において、ユーザ側が設定可能なデータの選択条件は以下である。

- 通常データ数
- 平日、休日別
- 曜日指定
- 目的データ、通常データの時間帯
- 目的データ、通常データの時刻間

目的データは、2017 年 1 月 11 日の 8 時から 20 時までの、1 時間ごとの各スポットの重要度とした。通常モデルは 3 つのモデルを作成した。表 2 に、各通常モデルを構成する通常データの選択条件を示す。1 つ目の通常モデル A は、過去 2 ヶ月間の平日のデータの集合とした。通常モデル A と比較することにより、目的データが「平日」としてどのような変化をしているかを知ることを目的とする。2 つ目の通常モデル B は、過去 2 ヶ月間の水曜日のデータ

表 2: 通常モデルの条件

通常モデル	A	B	C
	過去 2 ヶ月	過去 2 ヶ月	1 月全体
通常データ数	40 日	8 日	19 日
平日/休日	平日	平日	平日
曜日指定	無	水曜日	無
時間帯	8~20 時	8~20 時	8~20 時
時刻間	1 時間	1 時間	1 時間

の集合とした。通常モデル B と比較することにより、目的データを「水曜日」として捉えた場合に、どのような変化をしているかを知ることを目的とする。3 つ目の通常モデル C は、1 月の平日全体のデータの集合とした。解析範囲の季節性を考慮した場合、同じ月に属する日は似たような特徴を持つと考えられる。よって、通常モデル C と比較することにより、目的データを「1 月中の 1 日」として捉えた場合に、どのような変化が見られるかを知ることを目的とする。図 6 に、通常モデル別の各スポットの重要度の変化の度合いを示す。縦軸がスポット番号、横軸が時間である。色が濃いほど通常モデルからの変化の度合いが大きいことを意味する。なお、変化の度合いは、同一時間、同一スポットの通常モデルから推定された確率密度分布における目的データの情報量と定義した。例えば、目的データが、10 時におけるスポット S_1 の重要度の場合、通常モデル A は、過去 2 ヶ月間の 10 時におけるスポット S_1 の重要度で構成される。

本研究では、例として、各通常モデルにおいて、異なる振る舞いが見られた 3 つのスポットに着目する。1 つ目に、12 時におけるスポット R_2 に着目する。スポット R_2 は食堂である。図 6 の各図において、緑色の枠で囲んだ。図 6 にみられるように、過去の水曜日と比較した場合、変化の

度合いが大きいものに対し、他の2モデルでは、それほど変化が見られなかった。図7(a)に、12時におけるスポット R_2 の各通常モデルの確率密度分布と目的データの位置を示す。縦軸が確率密度、横軸がスポットの重要度である。また、凡例の σ は、ガウスクーネルのバンド幅である。通常の水曜日より、重要度が低く、あまり見られない事象であることが分かる。よって目的データは、平日としてはなく、水曜日としてみた場合に、その変化をとらえることが可能であると考えられる。2つ目に、14時におけるスポット L_1 に着目する。スポット L_1 は自習スペースである。図6の各図において、黄色の枠で囲んだ。図6にみられるように、各通常モデルにおいて変化の度合いが異なっている。図7(b)に、14時におけるスポット L_1 の各通常モデルの確率密度分布と目的データの位置を示す。過去の平日や水曜日からの変化の度合いが大きく、高い重要度を獲得している。それに対し、1月平日における変化は小さい。これは、1月は期末試験の時期でもあり、確率密度分布からもわかる通り、自習スペースの重要度が高い傾向は通常状態に近いことが原因であると考えられる。3つ目に、16時におけるスポット B_2 に着目する。図6の各図において、青色の枠で囲んだ。図6にみられるように、全ての通常モデルにおいて変化の度合いが大きい。図7(c)に、16時におけるスポット B_2 の各モデルの確率密度分布と目的データの位置を示す。これより、過去の平日や水曜日に比べ、バス停が通常より高い重要度を獲得していたことが分かる。また、1月の平日においても、例外的な重要度であった。4限目終了後の時間でもあり、帰宅するためにバスを利用する人が通常より集中したと推定される。図7の各図より、確率密度分布は異なった形状をしていることが分かる。よって、パラメトリックな推定手法より、カーネル密度推定を用いて個別に通常モデルを作成する方が実際の事象に適した結果が得られるといえる。

以上の結果から、通常モデルの適応的な作成による変化検出により、単一の比較対象を用いた場合では、変化を捉えることが困難な事象に対する変化検出が可能であることが示された。また、同一目的データの場合でも、異なる通常モデルと比較することにより、変化の割合に差が見られた。これは、目的の事象がどのような意味を持つのかを検討するための有益な情報となる。

5. まとめ

本研究では、グラフ理論における中心性を用いて求めたスポットの重要度に基づく、人流傾向の変化の適応的な検出手法を提案した。中心性は、各スポット間の移動人数とその流入関係も考慮に入れた SpotRank を提案した。SpotRank を用いることにより、実際の現象に合った重要スポットの検出が可能であることを示した。変化検出では、目的の事象と比較対象となる事象を、ユーザが柔軟に

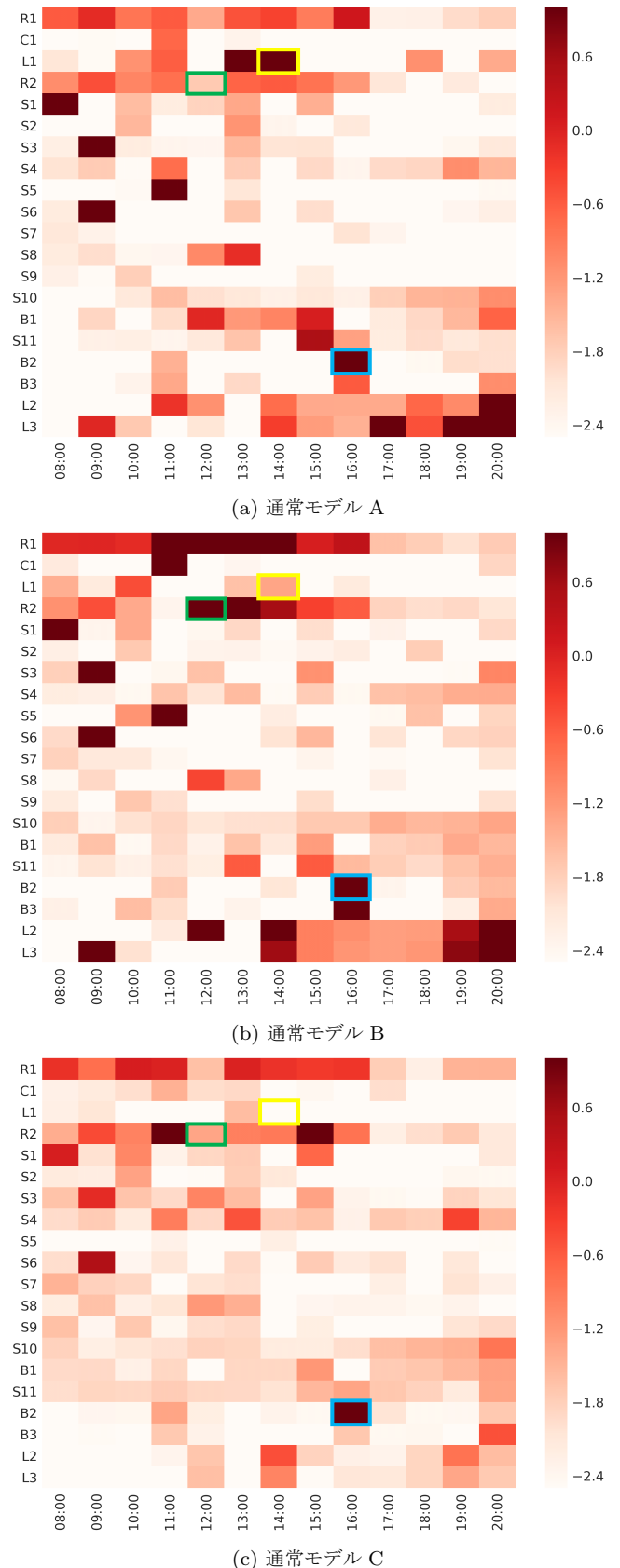
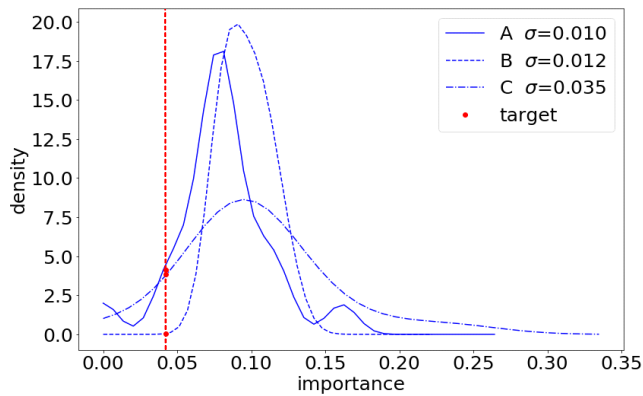
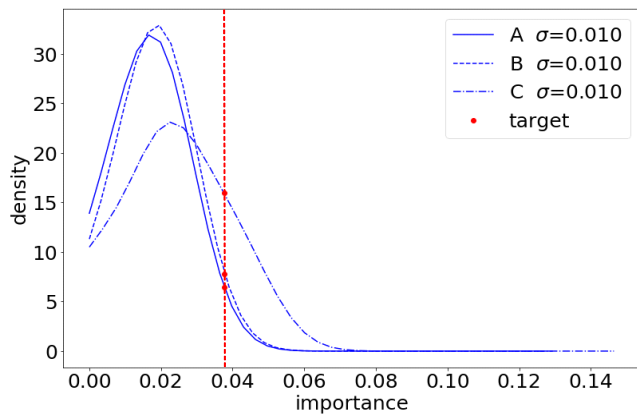


図6: 変化の度合い (2017/01/11 (水) 8~20時)

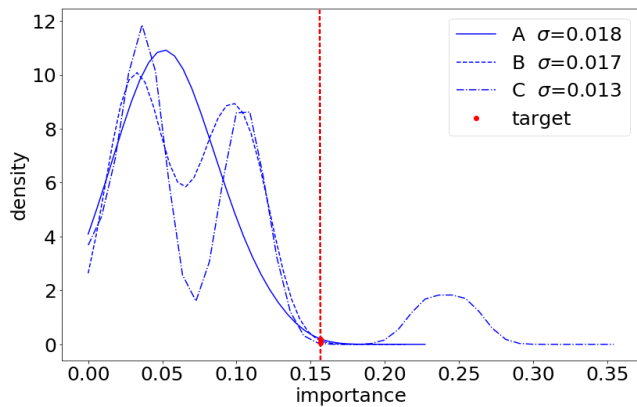
作成可能なシステムを作成した。通常モデルの確率密度分布は、カーネル密度推定を用いることで、各スポットでの異なる傾向の分布に対応した。その結果、通常モデルを複



(a) 12時, スポット R_2 (食堂)



(b) 14時, スポット L_1 (自習スペース)



(c) 16時, スポット B_2 (バス停)

図 7: 確率密度分布と目的データ (2017/01/11 (水))

数を用いて変化検出を行うことで、スポットがどのような性質を持った変化をしたのかを推定することが可能となった。

今後の研究では、実験場所については、より不特定多数の人々が訪れる、都市レベルでの解析や、ショッピングモール等の商業施設内での実験を行う。SpotRank については、移動人数に加え、スポットの特性も考慮に入れた中心性指標の改良を行う。また、変化検出については、Webブラウザ上で操作できるシステムの開発を行う。本研究では、通常モデルを作成する際の、データ選択条件をユーザー側で設定した。今後は、地図上での操作や、選択された目的デー

タに応じた、適切な通常モデルの自動作成など、ユーザーが容易に使用可能なシステムの開発を検討する。

謝辞

本研究の一部は、株式会社アドインテとの共同研究により実施された。

参考文献

- [1] 大西正輝, 依田育士. 大型複合施設における長期間にわたる人流比較と可視化手法. 電子情報通信学会論文誌 D, Vol. 93, No. 4, pp. 486–493, 2010.
- [2] 内田誠, 白山晋ほか. SNS のネットワーク構造の分析とモデル推定. 情報処理学会論文誌, Vol. 47, No. 9, pp. 2840–2349, 2006.
- [3] 松尾豊, 安田雪. SNS における関係形成原理-mixi のデータ分析-. 人工知能学会論文誌, Vol. 22, No. 5, pp. 531–541, 2007.
- [4] Chen Zhong, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, Vol. 28, No. 11, pp. 2178–2199, 2014.
- [5] Shuangming Zhao, Pengxiang Zhao, and Yunfan Cui. A network centrality measure framework for analyzing urban traffic flow: A case study of wuhan, china. *Physica A: Statistical Mechanics and its Applications*, Vol. 478, pp. 143–157, 2017.
- [6] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
- [7] Arthur S Alderson and Jason Beckfield. Power and position in the world city system. *American Journal of sociology*, Vol. 109, No. 4, pp. 811–851, 2004.
- [8] Hsin-Yu Shih. Network characteristics of drive tourism destinations: An application of network analysis in tourism. *Tourism Management*, Vol. 27, No. 5, pp. 1029–1039, 2006.
- [9] Karolina Taczanowska, Luis-Millán González, Xavier Garcia-Massó, Andreas Muhar, Christiane Brandenburg, and José-Luis Toca-Herrera. Evaluating the structure and use of hiking trails in recreational areas using a mixed gps tracking and graph theory approach. *Applied Geography*, Vol. 55, pp. 184–192, 2014.
- [10] Taras Agryzkov, Jose L Oliver, Leandro Tortosa, and Jose F Vicent. An algorithm for ranking the nodes of an urban network based on the concept of pagerank vector. *Applied Mathematics and Computation*, Vol. 219, No. 4, pp. 2186–2193, 2012.
- [11] 伏見卓恭, 斉藤和巳, 武藤伸明, 池田哲夫, 風間一洋. 実距離を考慮した中心性指標の提案と重要観光スポット抽出への応用. 人工知能学会論文誌, Vol. 30, No. 6, pp. 703–712, 2015.
- [12] 尾ノ上晃, 堀磨伊也, 島田敬士, 谷口倫一郎. 電波接触に基づく人々の多地点移動の可視化とパターン解析. 情報処理学会九州支部「火の国情報シンポジウム 2018」, 2018.
- [13] Amy N Langville and Carl D Meyer. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.
- [14] Simon J Sheather and Michael C Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 683–690, 1991.