

# グラフマイニングを用いたブロードバンドサービスエリア分類手法の一考察

Area classification consideration by graph-mining for broadband services

岩下 基\*<sup>1</sup>      高野 正次\*<sup>2</sup>      下川 信祐\*<sup>2</sup>  
Motoi Iwashita      Shouji Kouno      Shinsuke Shimogawa

\*<sup>1</sup>千葉工業大学経営情報科学科      \*<sup>2</sup>NTTサービスインテグレーション基盤研究所  
Chiba Institute of Technology      NTT Service Integration Labs.

In order to eliminate digital divide and promote regional development, it is important to install a fiber to the home (FTTH). It is essential to forecast potential demand in a target area for FTTH. Since the number of area which is managing telecom facility is large, effective area classification method is necessary as the preliminary step.

This paper proposes the efficient area classification method based on graph representation considering spatial characteristics of broadband services. We applied this method with broadband penetration rate for each area and to consider area characteristics in each group.

## 1. はじめに

ユーザ宅まで光ファイバを敷設（光化）した超高速通信網を構築することにより、様々なブロードバンドサービスが享受可能となり、情報格差の解消や地域振興の促進等が期待される。通信設備管理の最小単位となるエリアに対して、光化未整備エリアを光化することにより、光化整備率を向上することが望まれる。2009年度末で、全国の固定電話契約数に対して光を提供できる割合、即ち光化整備率は約84%と高い[1]。しかし、例えば、エリアに対応する市町村単位を見ると、2005年のデータではあるが、光化されていないエリアが特に人口規模の少ない地域で多いのが現状である[2],[3]。

光化の対象となるエリアは、一般にそのエリアの光サービスに対する潜在需要を推定することにより、判断するのが妥当である。特定のエリアに対して設備を投入するかの判断手法の観点から類似の検討として、対象となるエリアでのコンビニ等の店舗導入の意思決定に関する手法は、商圈分析の検討[4],[5],[6],[7]に述べられている。しかし、全国には光化未整備エリアが多いため、潜在需要の算出及びエリアの選択と決定には、膨大な時間を要する。そのため、効率の良いエリアの選定方法が必要となる。全国の光化を効率的に進めるためには、多くの地域をグループ化し、各々のグループでの光化の進め方（例えば、グループAに属する地域では、どういった特徴のエリアを光化するといった考え方）を決めていければ望ましい。

一般にブロードバンドサービス普及率の高いエリアを中心にその周辺は比較的普及率が高く、遠くなるに従い低くなる傾向があることから[8]、ブロードバンドサービスの普及をネットワーク形態として捉えた地域分類が必要である。このように様々なネットワークの分析には、グラフ理論を利用した研究が盛んである。グラフトポロジーの研究では、化学薬品等にグラフを用いた構造の解明や購買履歴等のクラスタリング手法が提案されている[9],[10],[11],[12]。しかし、通信サービスにおいて、複数の隣接するエリア間の関係を考慮したエリア分類の検討は従来ない。

本稿では、エリア毎の潜在需要を推定する前段として、経済的かつ効率的に光化するエリアを選択するための一手法とし

て、エリア間の隣接関係を地理的要因として考慮し、光化エリアを中心に複数のエリアから構成される全国の多くの地域をグラフ表現により分類する手法を提案した。エリアのブロードバンド普及率をもとに本手法を適用した地域分類例を示し、同じ分類に属する地域における光化未整備エリアの特性を考察した。

## 2. 本検討の位置づけとサービスの特徴

現状約1,000程度[2]の光化未整備エリアを対象に、どのエリアから光化するかは、一般にそのエリアの光サービスに対する潜在需要を推定することにより判断するのが妥当である。この光サービスの潜在需要を推定するのに、例えば近くに存在する既に光化されたエリアの需要変化を適用した方法が考えられる。もう少し一般的に述べると、複数の光化されたエリアと光化未整備エリアを対象に、重回帰分析等により複数の光化されたエリアの特徴をモデル化し、それを光化未整備エリアに適用して潜在需要を推定するという方法が考えられる（図1）。

ステップ2において、重回帰分析により光化されたエリアの特徴を的確に把握するため、ステップ1において、どういった複数エリアで構成される地域を選択するかがポイントとなる。しかも、エリア数は約7,000あり、その中で1,000程度の光化未整備エリアが散在している。本検討では、このステップ1にあたる部分を効率的に分類する手法を構築することにより、方針を立てやすくすることを目的とする。

一般に、ブロードバンドサービス普及率の高いエリアを中心にその周辺は比較的普及率が高く、遠くなるに従い低くなる傾向があることから、ブロードバンドサービスの普及形態を考慮した地域分類が予め必要である。そのために隣接するエリア同士のつながりに着目する。

## 3. 地域のモデル化

各エリアをノード( $n \in N$ )とし、エリア間を結ぶ鉄道や幹線道路をリンク( $l \in L$ )としてグラフ表現( $G = (N, L)$ )する。図2に複数のエリアをまとめたある地域を表現した例を示す。光化エリアから遠くなると普及率が低くなることから、本稿では以下の5種類のエリアを対象にする。

：光化されたエリア、 $n_{core} \in N_{core}$ 、両端が $n_{core}$ からな

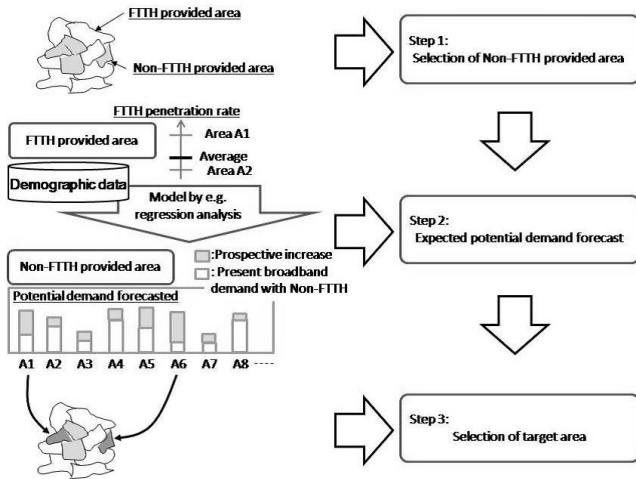


図 1: エリアの潜在需要算出フロー

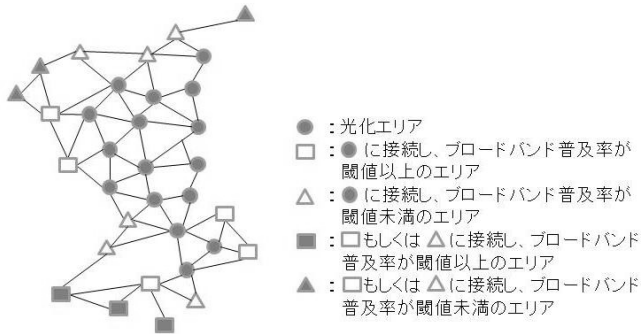


図 2: 地域のグラフ表現

るリンクの集合を  $L_{core}$  とし、 $G_{core} \equiv (N_{core}, L_{core}) \subset G$   
 :  $(n_{core})$  に隣接し、ブロードバンドサービスの普及率が  
 閾値以上の光化未整備エリア、 $n_{adj-1} \in N_{adj-1} \subset N$

:  $(n_{core})$  に隣接し、ブロードバンドサービスの普及率が  
 閾値未満の光化未整備エリア、 $n_{adj-2} \in N_{adj-2} \subset N$

:  $(n_{adj-1})$  もしくは  $(n_{adj-2})$  に隣接し、ブロードバン  
 ドサービスの普及率が閾値以上の光化未整備エリア  $n_{byd-1} \in$   
 $N_{byd-1} \subset N$

:  $(n_{adj-1})$  もしくは  $(n_{adj-2})$  に隣接し、ブロードバン  
 ドサービスの普及率が閾値未満の光化未整備エリア  $n_{byd-2} \in$   
 $N_{byd-2} \subset N$

ここで、 $N_{core} \cup N_{adj-1} \cup N_{adj-2} \cup N_{byd-1} \cup N_{byd-2} \in N$  と  
 なる。

#### 4. 分類手法

本章では、グラフ表現により特徴づけた分類アルゴリズム  
 を示す。グラフ表現により、 $G = (N, L)$  に対して、以下の分  
 類指標を計算する。

- $G_{core}$  のノード次数の平均 ( $d_G$ ) とその次数の平均からの  
 分散 ( $\sigma_G$ )

- $N_{adj-1}$  と  $N_{adj-2}$  それぞれから  $N_{core}$  への接続リンク数  
 の平均 ( $ln_{adj-1}, ln_{adj-2}$ )

- $N_{adj-1}$  と  $N_{adj-2}$  間のノード数比:  $r_1 =$   
 $|N_{adj-1}| / (|N_{adj-1}| + |N_{adj-2}|)$ , ここで  $|x|$  は、  
 “ $x$ ” の個数を意味する

- $N_{byd-1}$  と  $N_{byd-2}$  それぞれから  $N_{adj-1}$  への接続リンク  
 数の平均 ( $ln_{byd-1}, ln_{byd-2}$ )

- $N_{adj-1}$  と  $N_{byd-1}$  間のノード数比:  $r_2 = r_1 / (r_1 +$   
 $|N_{byd-1}| / (|N_{byd-1}| + |N_{byd-2}|))$

図 2 の例では、 $G_{core}$  の次数平均と分散は、それぞれ  $d_G =$   
 $3.5$  と  $\sigma_G = 1.87$  である。2 番目と 3 番目の指標は、 $G_{core}$  以  
 外のノードの  $G_{core}$  への連結度と普及率の関係を把握するた  
 めのものであり、図 2 の例では、各々  $ln_{adj-1} = 2, ln_{adj-2} =$   
 $1.67, r_1 = 0.45$  である。4 番目と 5 番目の指標は、普及率が  
 どこまで遠くに波及するかを把握するためのものであり、図 2  
 の例では、各々  $ln_{adj-1} = 1.33, ln_{adj-2} = 1.33, r_2 = 0.47$  と  
 なる。

$k$  個の地域をグラフ表現により、 $G^{(i)} \in S$  (ただし、 $i =$   
 $1, \dots, k$ ) とモデル化する。次数 ( $d$ ) とその分散 ( $\sigma$ )、 $r_1$  と  $r_2$   
 に関するノード数比 ( $\alpha$  と  $\beta$ ) を閾値として事前に設定し、以  
 下の手順でグラフの分類を実施する。

ステップ 1:

If  $d \leq d_{G^{(i)}} (i = 1, \dots, k)$  for  $\forall i$ ,  
 then  $G^{(i)} \rightarrow S_1$  and  $S - \{G^1\} \rightarrow S_2$

ステップ 2:

If  $\sigma \leq \sigma_{G^{(i)}}$  for  $\forall i$ ,  
 then if  $G^{(i)} \subset S_j$ ,  
 then  $G^{(i)} \rightarrow S_{j1}$  and  $S_j - \{G^{(i)}\} \rightarrow S_{j2} (j = 1, 2)$

ステップ 3:

If  $ln_{adj-1}^i \leq ln_{adj-2}^i$  and  $\alpha \leq r_1^i$  for  $\forall i$ ,  
 then if  $G^{(i)} \subset S_{jl}$ ,  
 then  $G^{(i)} \rightarrow S_{jl1}$  and  $S_{jl} - \{G^{(i)}\} \rightarrow S_{jl2} (j, l = 1, 2)$

ステップ 4:

If  $\beta \leq r_2^i$  for  $\forall i$ ,  
 then if  $G^{(i)} \subset S_{jlm}$ ,  
 then  $G^{(i)} \rightarrow S_{jlm1}$  and  $S_{jlm} - \{G^{(i)}\} \rightarrow S_{jlm2} (j, l, m = 1, 2)$

#### 5. 評価例と考察

図 3 に示す通り、光化された複数のエリアとその周辺エリア  
 からなる地域を全国から 7 ヶ所抽出し、ブロードバンド普及率  
 が与えられた閾値以上か未満かで光化未整備エリアを分類し、3  
 章で説明したグラフ表現を実施した。閾値として、 $C = 2.8, \sigma =$   
 $2.7, \alpha = 0.7, \beta = 0.5$  と設定した。分類に際して必要となる  
 各種分類指標を表 1 に示す。平均次数は、2.8~3.5 の間にあ  
 り、次数の分散は、1.6~2.7 の間に存在する。4 章で提案した  
 手法を適用した分類過程を図 4 に示す。ステップ 1 と 2 の次数  
 及び分散の閾値設定により分類は発生しないが、ステップ 3、  
 4 により分類された。

本稿では、先に述べたようにブロードバンドサービス普及率  
 の高いエリアを中心にその周辺は比較的普及率が高く、遠くな  
 るに従い低くなる傾向があることから、隣接エリア同士のつな  
 がりに着目し、グラフ表現を実施した。そのため得られた結果  
 が、対象とする地域内における人の交流状況や移動数等のデー

表 2: エリア内ノード情報統計

エリア	昼間人口		高齢世帯数	
	平均	分散平方和	平均	分散平方和
A	0.112	0.409	0.024	0.015
B	0.143	0.264	0.027	0.012
C	0.167	0.298	0.032	0.007
D	0.142	0.374	0.028	0.012
E	0.034	0.038	0.007	0.002
F	0.029	0.039	0.004	0.001
G	0.084	0.179	0.015	0.005

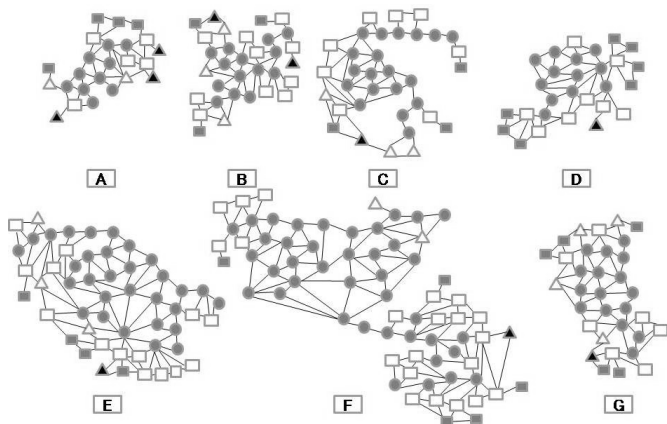


図 3: 評価対象地域のモデル化

表 1: 算出結果

エリア	$d_a$	$\sigma_a$	$n_{adj-1}$	$n_{adj-2}$	$n_{byd-1}$	$n_{byd-2}$	$r_1$	$r_2$
A	2.82	1.96	1.80	2.50	0.75	1.00	0.71	0.56
B	2.83	1.76	1.20	2.00	1.00	0.50	0.77	0.56
C	3.11	1.61	1.50	1.00	1.00	1.00	0.73	0.49
D	3.17	1.79	1.75	0.00	1.14	1.00	1.00	0.53
E	3.37	2.31	1.57	1.67	1.50	1.00	0.82	0.51
F	3.48	2.67	1.57	3.00	2.33	2.00	0.91	0.52
G	3.5	1.87	1.86	1.50	0.80	1.00	0.64	0.43

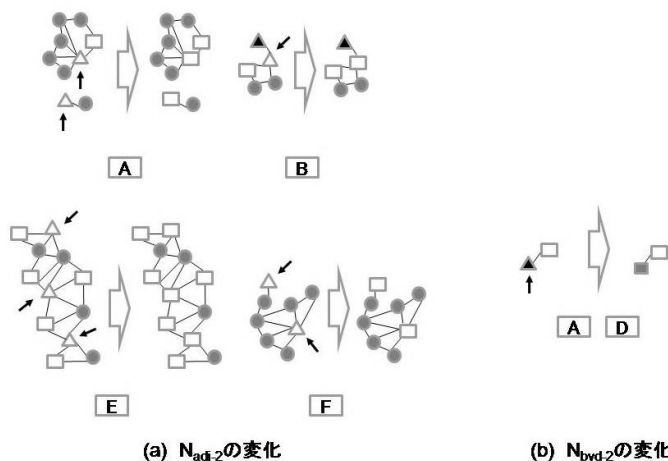


図 5: エリアの変化例

タと相関があるかを考察するのが望ましい。残念ながら地域内の交流状況に関する情報はほとんど存在しないが、ここでは総務省が発表している 2005 年の国勢調査データをもとに、各市町村の昼間人口（流動性によるサービスの普及効果を期待する項目）、高齢世帯数（非流動性の側面を考慮した項目）と各地域での結果との相関関係を考察した。全てがあてはまるデータは今のところ発見されていないが、昼間人口と高齢世帯数では、E, F は同類と見ることができ、昼間人口のエリア内のばらつきを見ると B, C, D が、また高齢世帯数のエリア内のばらつきを見ると、C, G は同じであることがわかった。グラフ表現により分類した結果が、これら複数の観点を含んでいると推測されるが、詳細分析は今後の課題である。

また、分類された地域内のエリアがどのように変化したか（与えられたブロードバンド普及率の閾値を超えたかどうか）を翌年の結果と比較した（図 5）。その結果、ステップ 3 の分類において、得られた A, B, E, F の地域では、光化エリアに隣接しかつブロードバンド普及率の低い光化未整備エリア ( $n_{adj-2}$ ) の普及率が高くなる特徴があることがわかった（図 5 (a)）。また、ステップ 4 において、A と D では、 $r_2$  の値が大きいので、光化未整備エリアに隣接する普及率の低いエリア ( $n_{byd-2}$ ) でも、翌年には普及率が高くなる特徴があることがわかった（図 5 (b)）。

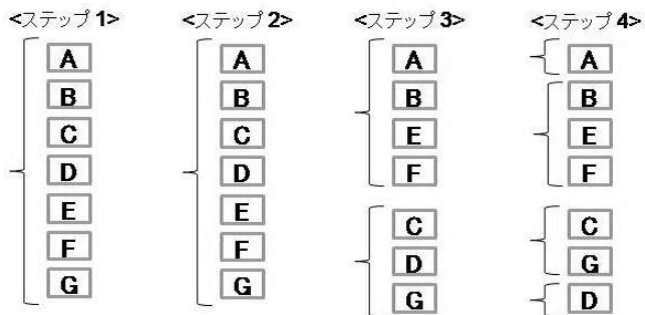


図 4: 地域分類結果

## 6. おわりに

本稿では、エリア毎の潜在需要を推定する前段として、経済的かつ効率的に光化するエリアを選択するための一手法として、エリア間の隣接関係を地理的要因として考慮し、光化エリアを中心に複数のエリアから構成される全国の多くの地域をグラフ表現により分類する手法を紹介した。エリアのブロードバンド普及率をもとに本手法を適用した地域分類例を示し、同じ分類に属する地域における光化未整備エリアの特性を考察した。地域内の交流状況との相関関係の詳細を見出すこと、各種要因となる係数を変化させ、普遍的な特性を発見していくことが今後の課題である。

## 参考文献

- [1] 情報通信白書 2009 年版, 総務省, (2009).
- [2] 全国均衡のあるブロードバンド基盤の整備に関する研究会 (2005-b) 「次世代ブロードバンド構想 2010 ディバイド・ゼロ・フロントランナー日本への道標」最終報告, 総務省, 2005 年 7 月, (2005).
- [3] 宿南: ブロードバンドサービスにおけるデジタルデバイドの解消政策について, 慶応義塾大学 メディアコミュニケーション研究所紀要, No. 56, pp.143-156, (2006).
- [4] 米田: エリアマーケティングの事実, 日経文庫, (2008).
- [5] 村山, 柴崎: G I S 理論, 朝倉書店, (2008).
- [6] 酒井他: Web-GIS による商圈分析ツールの構築および地域利用に関する実証的研究, 地球環境研究, Vol. 6, pp. 125-130, (2004).
- [7] 田中: 商圈構造の G I S 表現に関する検討, 日本計算機統計学会大会論文集, Vol. 17, pp. 139-142 (2003).
- [8] S. Shimogawa et. al.: "Mechanism of Diffusion of Fixed-Line Broadband Access Services in Japan and its Application to Long-Term Growth Prediction, -Rapid Diffusion in Urban Areas-", the 35th Annual Telecommunications Policy Research Conference, (2007).
- [9] 庄田他: 構造的類似性に基づくグラフクラスタリング, 第 17 回日本人工知能学会全国大会予稿集, 3F1-02, (2003).
- [10] 和田他: 部分構造情報を用いたグラフクラスタリング手法の検討, 第 20 回日本人工知能学会全国大会予稿集, 1A2-02, (2006).
- [11] 稲積他: グラフマイニングを応用した系列データ解析, 第 22 回日本人工知能学会全国大会予稿集, 2B3-02, (2008).
- [12] 高橋他: 化学物質の構造類似性にもとづくデータマイニング, J. Comput. Chem. Jpn., Vol. 2, No. 4, pp. 119-126, (2003).