

地区の持続可能性を高めるための 世帯マイクロデータを用いたデータマイニング

村中 智哉¹・宮本 和明²・大谷 紀子³・奥村 令子⁴

¹正会員 株式会社サーベイリサーチセンター 事業本部九州事務所
(〒812-0013 福岡県福岡市博多区博多駅東2-6-26)
E-mail: muranaka_t@surece.co.jp

²フェロー 東京都市大学教授 都市生活学部 (〒158-8586 東京都世田谷区等々力8-9-18)
E-mail: miyamoto@tcu.ac.jp

³正会員 東京都市大学教授 メディア情報学部 (〒224-8551 神奈川県横浜市都筑区牛久保西3-3-1)
E-mail: otani@tcu.ac.jp

⁴非会員 東急総合研究所 (〒150-0043 東京都渋谷区道玄坂1-10-7 五島育英会ビル2階)
E-mail: ryoko-o@triinc.co.jp

わが国における少子高齢化の波は世界的にも極めて急激なものである。これは都市自体が高齢化することを意味しており、様々な都市問題を誘発することが強く懸念されている。そのため、問題が顕在化するまでに適切な対応を取ることは、地方公共団体だけではなく関係するステークホルダーにとって重要な課題と言える。効率的な公共施策選択や公益企業等による公共的事業立案には、広い意味での公共サービス需要分析が必要である。そこで、本研究においては、世帯マイクロデータに基づくデータマイニングを通して、公共の施策立案あるいは公益企業の公共的事業立案等を支援する方法の構築を目的としている。まず、東急田園都市線沿線をケーススタディ地区として、入居時からの世帯変遷等の世帯マイクロデータと各種サービスに対する満足度や転居意志等に関するアンケート調査を実施した。その回答をもとにデータマイニング手法の一つである決定木手法を用いて、世帯マイクロデータと各種サービスに対する満足度や転居意志の関連を分析した。その結果、データマイニング手法を通して各種施策や対策の検討とそのターゲットセグメントを見いだす可能性を示した。

Key Words: Household Micro-Data, Household Segmentation, Decision Tree, Sustainable City

1. はじめに

わが国における少子高齢化の波は世界的に見ても極めて急激であり、また先進国の中でもいち早くこの問題に直面している。これは都市の高齢化を意味しており、適切な対策をとらなければ、様々な都市問題を誘発することが考えられる。そのため、問題が顕在化するまでに適切な対応を取ることは、地方公共団体だけではなく関係する公益企業をはじめとするステークホルダーにとって重要な課題と言える。効率的な公共施策選択や公益企業等による公共的事業立案には、広い意味での公共サービス需要分析が必要である。

そこで、本研究においては、公共の施策立案あるいは公益企業の公共的事業立案等を支援するために、詳細な世帯属性データに基づくデータマイニングを通して、公共的サービスごとに対象となる世帯をセ

グメントする方法の構築を目的としている。

都市における各種サービスに対する満足度は、同じサービスであったとしても、個々の世帯状況により大きく異なるものである。本研究では個々の世帯の詳細属性を世帯マイクロデータと呼び、それに基づく分析を行うものである。

研究グループでは、既に世帯マイクロデータを用いた政策手段の選択方法として、世帯をセグメント分割する方法を提案し、富山市を対象地域に適用している¹⁾²⁾。この研究においては、各種政策手段に対してのターゲットセグメントをデータマイニング手法により抽出している。また、この研究では、図-1に示すように、対象地域内の将来の世帯マイクロデータの分布の推計方法も構築している³⁾。これらを組み合わせることにより、将来の各種公共的サービスに対する需要を予測し施策選択を行う方法論を提案している。本研究ではこのアプローチを大都市

近郊のニュータウン地域に適用するための第一段階である。

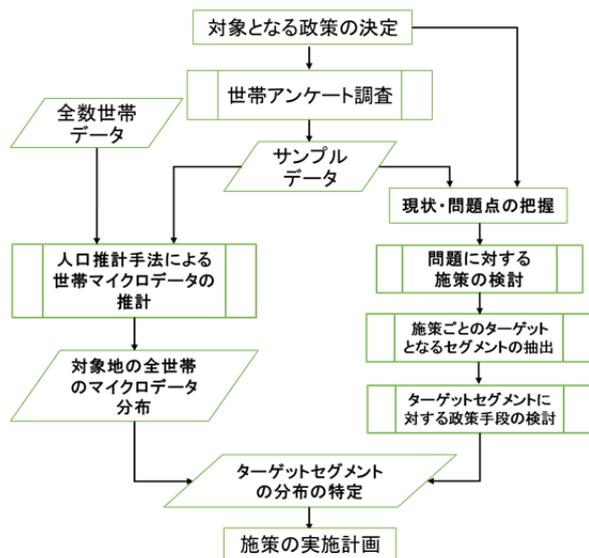


図-1 世帯マイクロデータに基づく施策選択³⁾

そこで本稿では、本システムを他の都市圏に適用する最初の試みとして、田園都市線沿線の一部を対象としたデータマイニングを実施し、都市政策のみならず、各種公共的施策選択に対してのデータマイニング手法の有効性を検討するものである。

2. 公共分野におけるマーケティング

(1) 概要

従来の都市政策は、一般に、対象地域内において画一的に実施されてきた。しかし、同じ政策に対しても目的通りの反応を示してくれる、政策目的から見て「好ましい反応」を示す世帯、あるいは「反応しない」世帯、逆に「好ましくない反応」を示す世帯も存在する。

それらに着目する上で有効的なのが、民間型のマーケティング手法である。民間型のマーケティング手法は、数多く存在する母数の中から自分たちのターゲットとなる対象集団、すなわちセグメントを見つけ出すために利用されているが、研究グループでは、これを応用し施策ごとのターゲットとなる世帯を抽出する方法について研究を進めている。

(2) 公共分野におけるマーケティング

これまで、公共分野あまり使われることがなかった民間型のマーケティングが、近年徐々に取り入れられてきている。海外においては、バスの乗車率向上⁴⁾、公共交通の乗車率向上⁵⁾、上水道インフラ事業⁶⁾や、人口計画⁷⁾等に用いられている。一方で日本国内での事例は、大阪府⁸⁾や青森県⁹⁾で実施された自治体住民に対するマーケティングの適用であったり、国が実施している「着地型観光」誘発のため

の「地域観光マーケティング促進マニュアル」¹⁰⁾などがあげられる。

しかし、本研究内で用いるような世帯の詳細データを用いた分析については見当たらない。

3. データマイニング手法

(1) データマイニング手法の選択

データマイニングの手法としては、統計的手法のほか、決定木や相関ルール、ニューラルネットワーク、サポートベクターマシンなどの機械学習技術が挙げられる。効果的なデータマイニングを実施するためには、データの形式や構造、目的に応じて、適切な手法を選択する必要がある。

本研究におけるデータマイニングでは、個々の世帯のマイクロデータから政策分析の対象となる世帯の特徴を抽出することを目的としている。すなわち、ターゲットセグメントを抽出するための規則を生成することが目的であるといえる。得られた特徴に基づいて政策を検討することを勘案すると、ニューラルネットワークのように処理過程がブラックボックスである手法ではなく、分類の基準が明確に表現される決定木のような分類手法が適している。

(2) 決定木

決定木とは、分類規則の木構造による表現技法である。分類対象となる事例は、複数の属性と各属性に相当する属性値、および事例が属するクラスとなり、決定木は属するクラスが既知の事例を用いて生成される。

決定木の例を図-3に示す。決定木の非終端ノードには属性、アークのラベルには属性値に関する条件、終端ノードには分類結果となるクラスが割り当てられる。1つの非終端ノードと、その下に結合しているアークの組によって1つの分類条件を表現する。事例の各属性値に従って根ノードから終端ノードまで決定木をたどることで、事例の属するクラスが判定される。根ノードから終端ノードまでの経路が1つの分類規則を表し、非終端ノードとアークによる分類条件の連言が前件、終端ノードが後件となる。

(3) 決定木生成手法

決定木生成に関するさまざまなアルゴリズムが提案されており、代表的なシステムとしてはCART¹¹⁾、ID3¹²⁾、C4.5¹³⁾などが挙げられる。これらのシステムでは、分類誤り率による枝刈りやブースティングアルゴリズム等の利用により、未知事例分類におい

て高い正解率を得ている。しかし、正解率の高い決定木は複雑でノード数が多いという傾向が一般に知られている。ノード数は決定木の表す分類規則の解釈の容易さを大きく左右する。

進化計算アルゴリズムの一種である共生進化を用いて決定木を生成する手法が提案されており、C4.5よりも正解率が高く、簡素な決定木が生成できることが示されている¹⁴⁾。本研究では、本手法に基づき決定木を生成する。

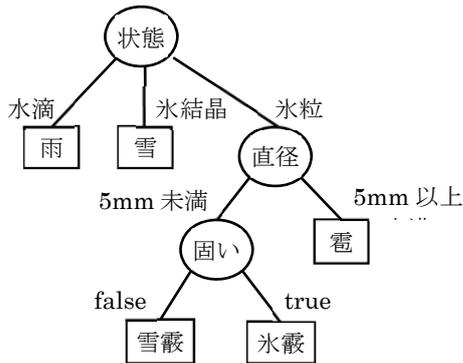


図-2 決定木の例

4. アンケート調査と分析手順

(1) アンケート調査の実施概要

利用するデータは田園都市線沿線で新たに実施したアンケート調査の回答結果である。アンケート調査の実施概要は以下の通りである。

- ・実施期間：2014年12月から2015年1月
- ・対象地域：多摩田園都市第3ブロック
- ・対象世帯：全51,601世帯から30%の無作為抽出
- ・送付票数：14,500票
- ・回収数：1,572票(10.8%)
- ・回収方法：返信用封筒を同封しポストイング、郵送回収

配布対象の田園都市第3ブロックは東急田園都市線の藤が丘、田奈、青葉台の各駅周辺の地域である。配布対象地域の地図を図-3に示す。

(2) 調査内容

質問の項目としては、研究グループが以前実施した富士市でのプロジェクトでの知見を反映して以下のように設定している。

- ・世帯属性等：世帯構成員の個別属性と現住居への入居時の世帯構成と各属性
- ・現在の住居の詳細：住居の形態や地区年数等、居住地の満足度と入居時からの比較
- ・交通や買い物：最寄り駅や自動車、自転車等の保

有台数

- ・転居前の住居：住居の形態や最寄り駅、間取り等
- ・住居の今後：住み続けるにあたっての不安、住み替えの意思等



図-3 アンケート調査対象地域

これらの項目をアンケートに取り入れ、特定されたターゲットセグメントが、何に満足し、何に不安を持っているかを分析し、施策立案あるいは公共的事業立案等に活かす。

(3) 分析手順の設定

本研究でのデータマイニングの一連のプロセスを図-4に示す。このプロセスは2段階に分かれており、第1段階では定性的なデータのみを用いて施策立案あるいは公共的事業立案等のターゲットとなりえる「意思」を持つセグメントの発見を目指す。

第2段階では、定量的なデータ、つまり世帯属性を用いて先に発見した定性的なデータを持つ世帯がどのような属性を持っているのかを分析する。

ターゲット特定プロセス

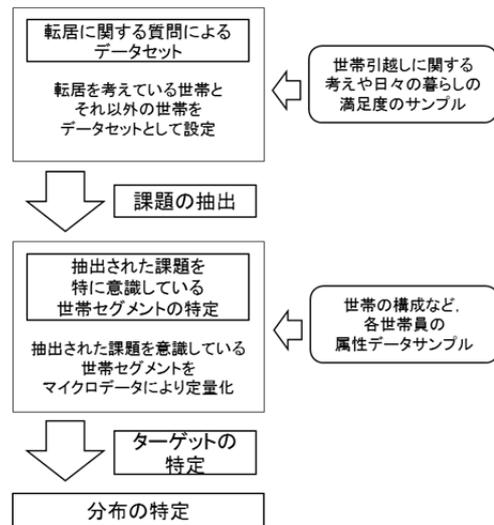


図-4 ターゲットセグメンテーションの手順

4. 決定木による分析結果

(1) 第1段階の分析結果

第1段階では、転居意思、特に対象地域外に転居意を考えている世帯を抽出し、それらの世帯が現在・未来において「何に困っているのか」「何が不安なのか」を分析し、第2段階での世帯属性の割り出しへと繋げる。

また、転居の「しやすさ」に1番関わると考えられる、住居の所有形態について分割して分析を実施する。この所有形態は持ち家か賃貸かで分類する。社宅や寮は賃貸に分類するものとする。

a) 賃貸住宅

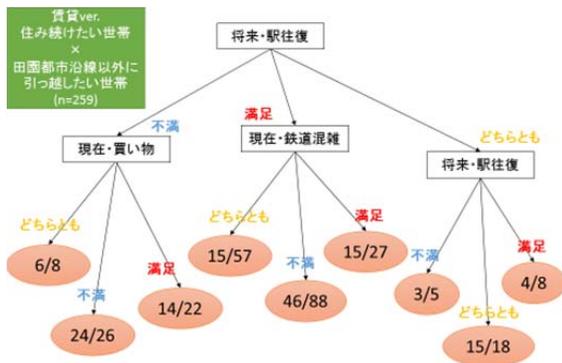


図-5 賃貸住宅の第1段階の決定木

賃貸住宅では、主に現在の買い物、また将来の駅の往復に不安を抱いていることがわかる。また、将来の駅への往復に不安を抱いている世帯が少ないのは、持ち家に比べて転居がしやすいということが根底にあると考えられる。つまり、将来の不安自体はあまり重要ではなく、現在の物事の方を重要視し易い傾向にあることがわかる。

b) 持ち家

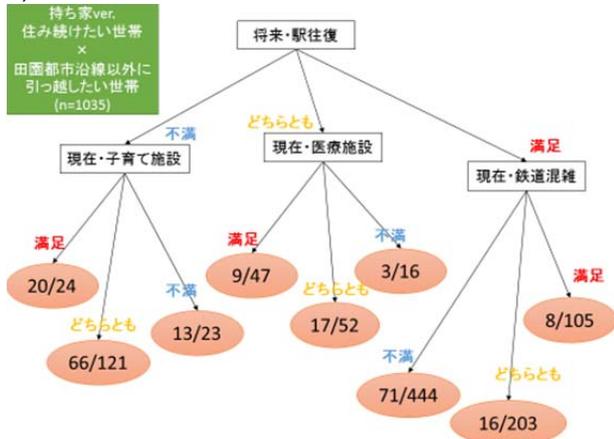


図-6 持ち家の第1段階の決定木

持ち家では、子育て施設や医療施設、将来の駅の往復に不満を抱いていることがわかる。一方で、将来の駅への往復に不安はないが、現在の鉄道の混雑に不満を抱いている世帯が一定数いることがわかる。

(2) 第2段階の分析結果

第1段階で、得られた結果から第2段階の分析を実施する。使用するデータは前項で賃貸・持ち家それぞれの分割で特定されたものである。

これらを世帯属性に基づいてデータマイニングを実施し、それぞれの項目について分析を実施する。

a) 賃貸住宅

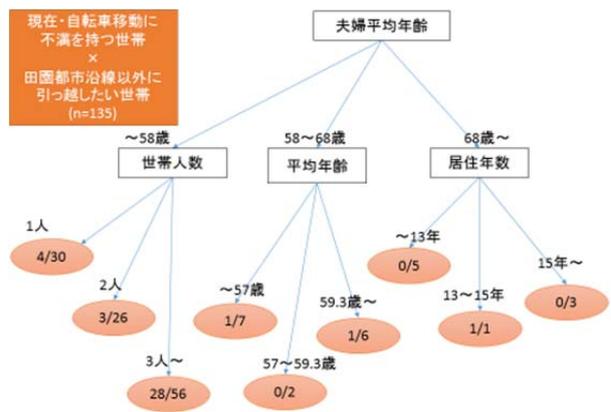


図-7 賃貸住宅世帯の第2段階の決定木 (将来の自転車利用への不安を持つ世帯)

図-7は賃貸住宅に住む世帯で、自転車での移動に不満を持つ世帯を探す木である。この木では夫婦の平均年齢が58歳以下で、世帯の人数が3人以上の世帯が特に不満を抱いていることがわかる。

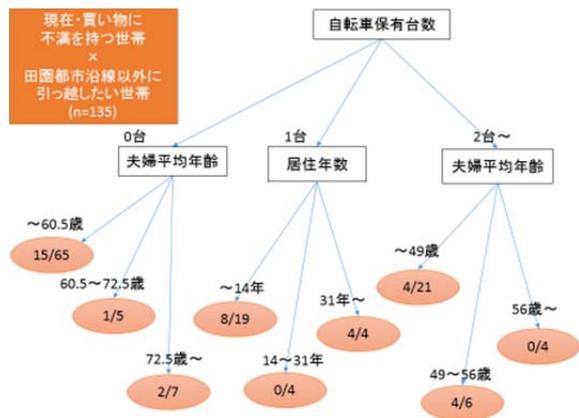


図-8 賃貸住宅世帯の第2段階の決定木 (現在の買い物に不満を持つ世帯)

図-8は賃貸住宅に住む世帯で、買い物に不満を持つ世帯を探す木である。

b) 持ち家

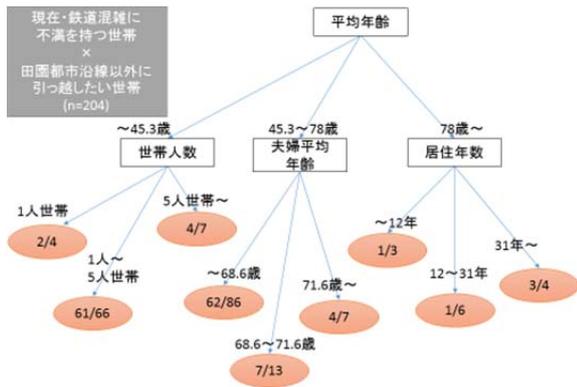


図-9 持ち家世帯の第2段階の決定木 (現在の鉄道混雑に不満を持つ世帯)

図-9は持ち家に住む世帯で、鉄道の混雑に不満を持つ世帯を探す木である。

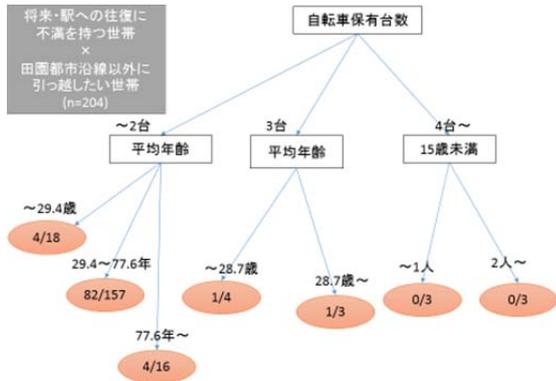


図-10 持ち家世帯の第2段階の決定木 (将来の駅への往復に不安を持つ世帯)

図-10は持ち家に住む世帯で、将来駅への往復に不安を持つ世帯を探す木である。

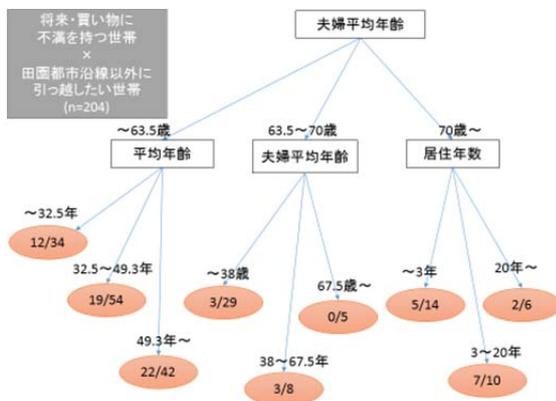


図-11 持ち家世帯の第2段階の決定木 (将来の買い物に対する不安を持つ世帯)

図-13は持ち家に住む世帯で、将来の買い物に不安を持つ世帯を探す木である。

(3) 段階を設定しない分析結果

ここまで、いくつかの結果を紹介してきたが決定木の葉の部分にあたる、終端ノードの一つ一つの母数があまり大きくないため結果に不安が残る。

しかし、しっかりと結果を出している木もあることから、データセットの作成と分析に利用する属性の取捨選択の重要性がわかる。安定して結果を出すことができれば住民サービス需要分析の一つの手段として確立されるだろう。

ここまでは住居の形態に基づいて分類し分析を進めてきた。しかし、この方法では母数が少なくなってしまい結果の説明力が薄れてしまう可能性がある。そこで住居分類を行わないケースでの分析結果を図-12から図-14に示す。

a) 転居希望世帯の決定木

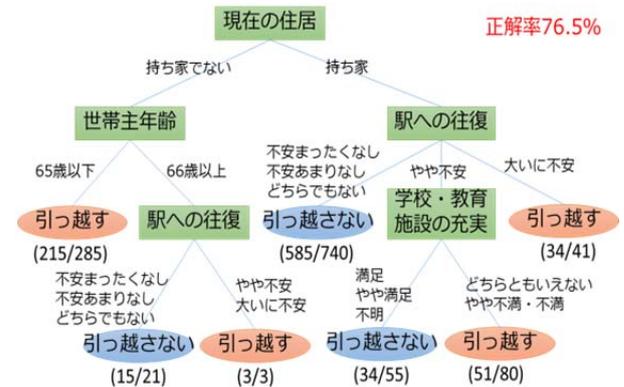


図-12 転居希望世帯の決定木

図-12 に示す決定木は、現在の住居が持ち家でなく世帯主が65歳以下の世帯、現在の住居が持ち家でなく世帯主が66歳以上で駅への往復に多少なりとも不安を感じている世帯、現在の住居が持ち家で駅への往復に大いに不安を感じている世帯、現在の住居が持ち家で駅への往復にやや不安を感じ、学校・教育施設の充実度に不満を感じている世帯が転居を希望していることを表している。

決定木の終端ノードに到達するデータ数を考慮すると、現在の住居が持ち家でなく世帯主が65歳以下の世帯は引っ越しを望み、現在の住居が持ち家で駅への往復に不安を感じていない世帯は引っ越しを望んでいない、という規則が大勢を占めているといえる。

また、転居の項目以外に着目した木が以下のものである。

b) 将来の駅への送迎に不安を持つ世帯の決定木

図-13 に将来の駅への送迎に不安を持つ世帯の決定木を示す。

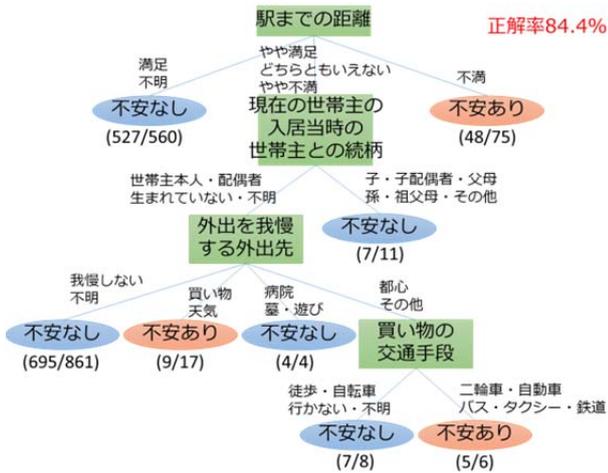


図-13 将来の駅への送迎に不安を持つ世帯の決定木

図-13 に示す決定木は、駅までの距離に不満を感じている世帯、駅までの距離に満足でも不満でもなく、現在の世帯主が入居当時は世帯主、世帯主の配偶者、未誕生であり、天気が悪いときの外出や買い物を我慢する世帯、駅までの距離に満足でも不満でもなく、現在の世帯主が入居当時は世帯主、世帯主の配偶者、未誕生であり、都心への外出を我慢し、買い物には二輪車、自動車、バス、タクシー、鉄道を利用する世帯が、将来の駅への送迎に不安を感じていることを表している。

決定木の終端ノードに到達するデータ数を考慮すると、駅までの距離に満足している世帯や、外出を我慢しない世帯は、駅への送迎に不安を感じていない、という規則が大勢を占めているといえる。

c) 将来の買い物に不安がある世帯の決定木

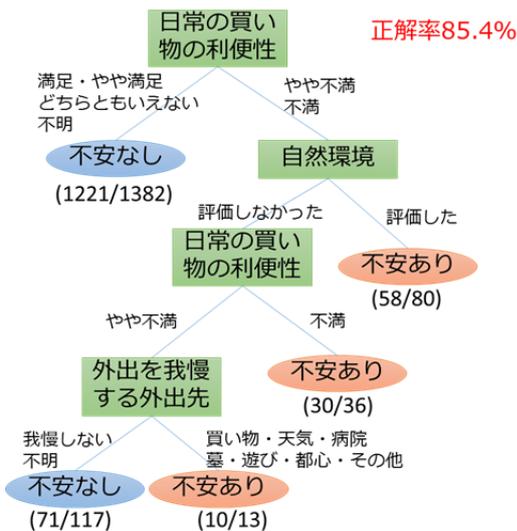


図-14 将来の買い物に不安がある世帯の決定木

図-14 の決定木は将来の買い物に不安がある世帯対するものである。この木は日常の買い物の利便性に不満を感じていない世帯、日常の買い物の利便性にやや不満を感じており、現在の住居を決める際に自然環境を評価せず、外出を我慢しない世帯は、買い物に不安を感じていないことを表している。大勢を占めているのは、現状での買い物の利便性に不満を感じていない世帯は買い物に不安を感じていないという規則である。

5. おわりに

本研究では、公共の施策立案あるいは公益企業の公共的事業立案等を支援するために、詳細な世帯属性データに基づくデータマイニングを通して、公共的サービスごとに対象となる世帯をセグメントする方法の構築を行った。

本研究の成果としては、以下のことがあげられる。まず、アンケート調査により世帯のマイクロデータと現在における満足度と将来に対する不安、さらに転居意志に関する基礎的データを入手した。それを用いて各種サービスに対して不満を持つ世帯や将来に不安を持つ世帯をセグメントする方法を作成し、データを用いて分析を行った。分析結果に関しては今後の考察が必要ではあるが、公共機関あるいは公益企業が今後の施策や方向性を検討するための基礎的資料となることのある程度確認できた。残された課題はあるものの広い意味での公共サービス需要の分析のための一つの手段として世帯マイクロデータを用いたデータマイニングの手法の有効性がある程度示されたと考えている。

なお、アンケート対象地域は首都圏のニュータウンの中でも一般的に特に優良と評価されている地域であることから、明確な課題を抽出するケーススタディとしては必ずしも適した地域とは言えない。そのため、特徴的な課題抽出にまでは至っていないと言える。

また、既に説明しているように、研究グループ内では、国勢調査等の全数調査と対象地域から収集される一部の世帯の完全な世帯マイクロデータから、対象地域全ての世帯、また世帯に所属する個人の「属性」を推計するエージェントベースの人口推計手法を開発している。これにより対象地域に存在する全ての世帯、個人を対象としての各種分析が可能となっている。

このデータとの組み合わせにより、ターゲットとなる世帯がどこに多く分布しているかを把握することが可能となるため、より効率的な公共の施策立案あるいは公益企業の公共的事業立案等を支援することができると思われる。

謝辞：アンケートの実施に際しては、複雑な質問内容にもかかわらず、対象地域の多くの皆様からご回答をいただいた。記して謝意を表する次第である。

参考文献

- 1) 福岡裕介, 宮本和明, 北詰恵一, 鈴木温: 富山市における住宅立地と交通に関する実態調査, 第 47 回土木計画学研究発表会論文集, CD-ROM, 2014.
- 2) 福岡裕介, 大谷紀子, 杉木直, 宮本和明: 世帯マイクロデータに基づく都市政策手段の選択方法, 第 49 回土木計画学研究発表会論文集, CD-ROM, 2014.
- 3) 村中智哉, 杉木直, 福岡裕介, 宮本和明: 世帯属性に基づく潜在住み替え需要の分析と政策手段の検討, 第 50 回土木計画学研究発表会論文集, CD-ROM, 2014.
- 4) Cyrus Njiru, Kevin Sanson: Strategic Marketing of urban water services, WEDC, 2001
<http://wedc.lboro.ac.uk/resources/conference/27/Njiru.pdf>
- 5) TRB, Transportation Research Board: A Handbook: Using Market Segmentation To Increase Transit Ridership, TRCP Report 36, 1998
- 6) Pierre Laconte: Smart Segments For Urban Public Transportation: An International Survey of Practices, Japan Railway & Transport Review 32, 2002, pp.11
http://www.jrtr.net/jrtr32/pdf/f04_lac.pdf
- 7) Suneeta Sharma, William Winfrey, Mona Marin: A Family Planning Market Segmentation Analysis in Romania, POLICY, 2001
http://www.policyproject.com/pubs/countryreports/Rom_MarkSeg.pdf
- 8) 大阪府: 自治の窓 持続可能な行政経営について 民間経営手法の更なる活用,
<http://www.pref.osaka.lg.jp/shichoson/jichi/jichimado4.html>
- 9) 上山信一, 玉村雅敏: 先進性秘めた青森県の政策マーケティング手法, 地方行政, 2001.
- 10) 国土交通省総合政策局旅行振興課: 地域観光マーケティング促進マニュアル, 2006.
- 11) Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J.: Classification and Regression Trees, Wadsworth & Brooks, 1984.
- 12) Quinlan, J.R.: Induction of Decision Trees, Machine Learning, Vol.1, No.1, pp.139-159, 1986.
- 13) Quinlan, J.R.: C4.5: Programs for Machine Learning, Morgan Kaufmann, 1993.
- 14) 大谷紀子, 志村正道: 共生進化に基づく簡素な決定木の生成, 人工知能学会論文誌, Vol.19, No.5, pp.399-404, 2004.

(2015.4. 24 受付)