

パーソントリップデータからの確率的潜在時空間意味構造モデリング ～地域活性化や観光サービスへの応用を目指して～

Probabilistic Latent Spatiotemporal Semantic Structure Models Based on Travel History Data

廣川典昭^{*1*2} 村山敬祐^{*2} 本村陽一^{*1*2}
HIROKAWA Noriaki MURAYAMA Keisuke MOTONMURA Yoichi

^{*1}産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

^{*2}東京工業大学大学院
Tokyo Institute of Technology

Countermeasures for increasing vacant stores in the rural area are considered to play an important role for regional revitalization. The utilization of the physical and psychological closeness of local shopping districts can suggest a solution in the future. Therefore we suggest information recommendation in real time. In this paper specifically we estimate the context of the travels by bayesian network to recommend information related to the context in real time.

1. はじめに

近年になっても地域商店街での空き店舗の増加はあとを絶たず、地域活性化における重要な問題の一つである。大型店舗と比較すると、地域商店は、物理的・心理的な距離が近く、その利便性を最大限に活用できれば、地域活性化への糸口となり得ると考えられることから、筆者らは、地域商店へ人々を促すためにリアルタイムに情報推薦を行う方法についての検討を行う。リアルタイムに情報推薦を行うためには、ユーザタイプ、ユーザの行動のコンテキスト（行動目的）、ユーザの現在状況という3つの要素を適切に推定する必要があると考えられる。Hirokawaらは、その中でも人々の行動目的を推定するために、人々の移動履歴を確率的潜在意味解析(pLSA)[Hoffman 1999]により分析し、地域と移動者の活動との関連性を潜在クラスとして抽出した[Hirokawa 2015](図1)。

本稿では、この抽出された地域特性と、時刻、性別等の移動者のデモグラフィック属性等をベイジアンネットワーク[PEARL 1985]で解析することにより、駅等への到着時点において行動目的の推定が可能であることを示す。

2. 人の流れデータについて

本研究は、東京大学空間情報科学研究センター(CSIS)の空間データ利用を伴う共同研究であり、CSISが作成した東京都市圏パーソントリップ調査に基づく「人の流れデータ」を用いて分析を行う。「人の流れデータ」は、東京都市圏における2008年10月1日0:00から同日23:59までの移動を記述した588568名分の移動履歴データである。1日の移動は、通勤や買い物等の単一の移動目的を持つ単位(トリップ)に分割されており、その各々に対して、1分毎の推定された位置情報(緯度・経度)に加えて、デモグラフィック属性(年齢・性別・居住地・職業)や、移動の目的、当該時刻で利用中の交通手段の情報が付与されている。

「人の流れデータ」を用いる理由は、アンケート結果に基づく移動履歴データである性質上、GPSや各種センサー等の他の移動履歴データでは得にくい移動の目的や利用した交通等の情報が入手できるためである。また、本稿では、データ

サイズ等の観点から、東京23区の「人の流れデータ」を用いて分析を行うが、pLSAで抽出した地域特性をベイジアンネットワークを用いて説明することのできる確率的潜在時空間意味構造モデルを用いれば、地方都市でも適用可能であることから地域活性化への一助となり得ると考えられる。

2.1 人の流れデータの基礎集計

図2に「人の流れデータ」を集計した値を示す。平日ということもあり、通勤・通学による移動が36.1%の割合で存在していることが確認された(図2(a))。また、公共交通機関を用いて比較的長距離を移動している人々と、徒歩や自

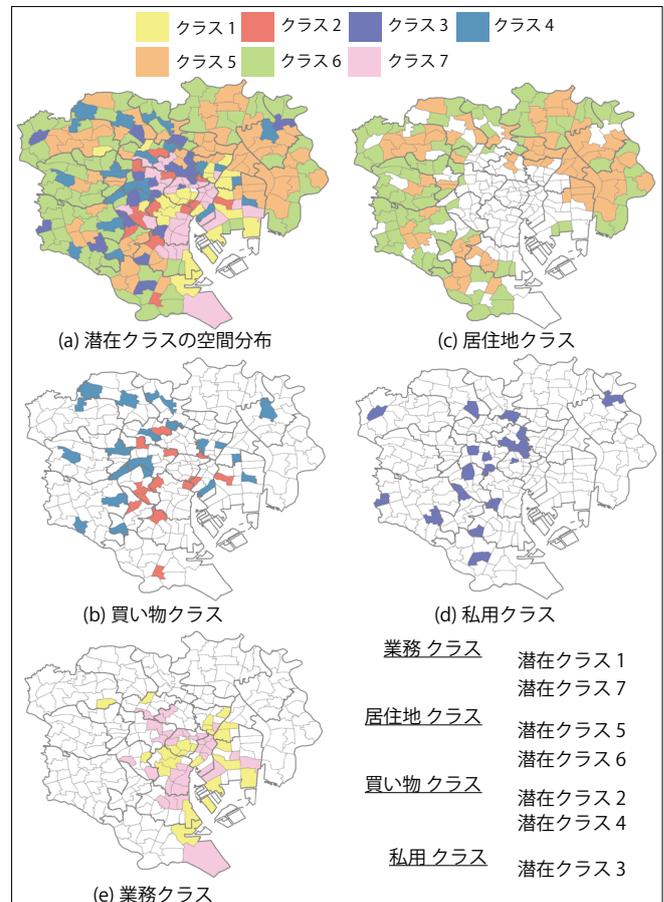


図1: 移動履歴から抽出された潜在クラスの空間分布

* 産業技術総合研究所
〒135-0064 東京都江東区青海 2-3-26
産業技術総合研究所 臨界副都心センター本館 331
E-mail: hirokawa.n@aist.go.jp

転車等で短時間の移動を行う人々が回答者の大半を占めることがわかる(図2(b)). 主夫・主婦や学生等は, 全体の約25%を程度であり, 75%以上が通勤等の拘束の多い職業従事者であることが確認できた(図2(c)). 地域活性化のためには, このような日常生活の中で適切な情報推薦を行う必要があると考えられることから, すべてのデータを対象として分析を行った.

3. ベイジアンネットワークによる移動目的の推定

3.1 行動目的推定のためのベイジアンネットワークの構築

本稿では, ベイジアンネットワークを用いて, 移動者の行動目的の推定を行う. モデル構築の際には, ベイジアンネットワーク構築ソフトウェアである Bayonet[本村 2003]を使用した. 具体的な推定手順を以下に示す.

手順1: 移動履歴データから無作為に4セットの学習データとテストデータを作成する.

手順2: 各々のデータセットについて, 学習データからベイジアンネットワークを構築する.

手順3: 手順2で構築したベイジアンネットワークを用いてテストデータの行動目的を推定し, ホールドアウト検証を行う.

分析対象とした移動履歴は, 23区内に目的地をもつ全389377トリップであり, このデータを無作為に5等分し, そのうち4つを学習用データ, 残りの1つをテストデータとする組を4セット作成し, ホールドアウト検証を行った. なお, ベイジアンネットワークに入力した変数は, 性別, 年齢区分ダミー(5区分), 時間帯ダミー(7区分), 所要時間ダミー(4区分), 職業ダミー(6区分), 主要交通手段ダミー(6区分), 目的地の潜在クラスダミー(7区分)である.

構築したベイジアンネットワークの行動目的を説明する部分を抽出し, 図3に示す. pLSAにより抽出された買い物クラスや居住地クラスが「買い物」「自宅へ」を説明していることが確認できる. また, 時間帯の変数の種類が多く抽出されており, 通勤や日常的な買い物等は, 時間帯に大きく依存する移動目的であることがわかる. 徒歩・自転車による移動が買い物と, 業務関係以外を説明していないことも興味深い. 推定精度(表1)は, 確率が最大となる移動目的で推定した場合であっても, 全体で64.7%と高精度で推定できており, 確率が2位の行動目的まで含めると, 推定精度は, 81.6%にまで上昇することから, 本研究の目的の一つである駅等への到着時点での移動目的の推定が可能となることを示している.

3.3 推定精度と日中の活動多様性

前節では, 全体的な振る舞いとして, 80%以上の高精度で移動目的の推定が可能であることを示した. 本節では, さらにミクロな分析を行い, 推定精度を低下させている要因についての検討を行う. 図4に行動目的ごとの正誤割合を示す. 自宅へ, 買い物, 通勤・通学の移動目的は, どれも80%以上の高精度で推定できていることが確認できた. これらの移動目的は, トリップ全体に占める割合も多いことから(図2(a))このような高精度での推定が可能となったものと考えられる. しかし, 私用や業務関係の移動目的の推定精度は低く, 現状のままでは推定が困難な可能性がある. 精度の低い変数を確認すると, 主に主夫・主婦や営業等の日中に多様な移動を多く行う人々が原因である可能性が判明した.

4. まとめ

地域商店の最寄り駅等へ人々が到着した時点でリアルタイ

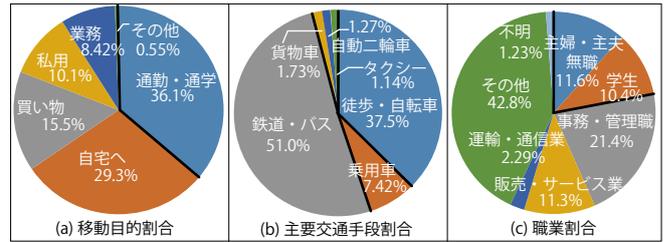


図2: 人の流れデータ単純集計

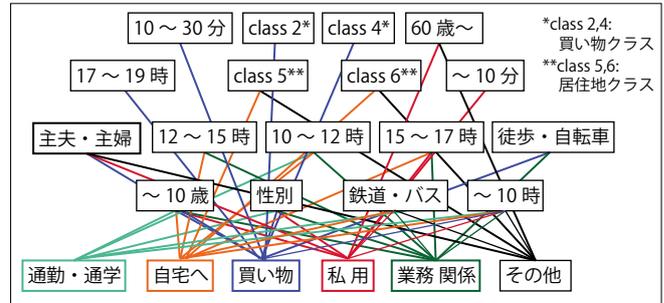


図3: 移動目的を説明するベイジアンネットワーク(抜粋)

表1: 各検証時の推定精度

精度 (%)	1回目	2回目	3回目	4回目	全体
最大確率で推定	64.5%	64.8%	64.7%	64.7%	64.7%
2位以内に入る	81.5%	81.5%	81.5%	81.5%	81.6%

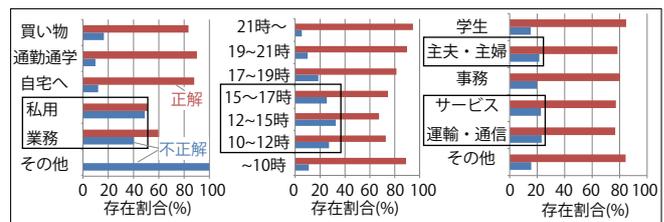


図4: 推定精度と日中の活動多様性(上位2位の精度)

ムに情報推薦を行うことを目的とし, 人々の移動履歴データの分析を行った. pLSAによって抽出された地域特性の潜在クラスや, 人々の移動履歴データをベイジアンネットワークを用いて解析することで, 全体で81.6%という高精度で行動目的の推定が行えることを示した. しかし, 目的別にみると, 移動履歴データの少ない私用目的の移動や, 業務関係の移動では, 推定精度が高くないことも判明した. これらは, 主夫・主婦やサービス業者等による日中の多様な移動が原因であると考えられる. これらを踏まえた推定精度の向上が, 今後の課題である.

参考文献

[Hoffman 1999] Hoffman, T.: Probabilistic Latent Semantic Analysis, Proc. UAI '99, pp. 289-296 (1999).
 [Hirokawa 15] Hirokawa Noriaki, Keisuke Murayama, Yoichi Motomura: Probabilistic Latent Spatiotemporal Semantic Structure Models Based on Travel History Data for Regional Revitalization, ICserv 2015 (2015)
 [PEARL 1985] PEARL, Judea. Bayesian networks: A model of self-activated memory for evidential reasoning, University of California (Los Angeles). Computer Science Department(1985).
 [本村陽一 03] 本村陽一: ベイジアンネットワークソフトウェア Bayonet, 計測と制御, vol. 42, No. 8, pp. 693-694 (2003)