

時空間行動データ検索システムの開発とその評価*

Development of Search System for Action Data between Space-time and the Evaluation *

高田普丈**・河野浩之***・仲野潤一****・羽藤英二*****

By Hirotake TAKATA**・Hiroyuki KAWANO***・Junichi NAKANO****・Eiji HATO*****

1. はじめに

移动通信システムを活用したプローブパーソン技術の発達により、緯度経度座標計測結果をベースに、人のトリップOD、移動・滞在時間あるいは経路など、詳細・大規模・高精度かつシームレスな時空間行動データの取得が可能となってきた。また、交通施策評価・整備効果把握に加え、避難訓練調査、観光調査、マーケティング調査等への新たな分野へのシステムに柔軟に適用されることによるWEBダイアリー機能の発達・標準化が行われつつある。しかしながら、緯度経度座標計測結果をベースとした調査データの評価・解析は多く報告されているが、WEBダイアリーデータを含んだ時空間行動データの検索技術が確立されているとは言いがたい。また、時空間行動データに紐付けられるWEBダイアリーデータの評価・解析は、人の行動実体、その行動が取られる原因・理由への把握をより詳細に行える可能性がある。

そこで本開発では、プローブパーソンシステムで得られた人の活動・行動変動に対する時空間行動データの評価・解析への足がかりとして、時空間行動データと、WEBダイアリーデータの並列化による、時空間行動データ検索システムの開発とその評価を行った。

*キーワード：プローブパーソン、システム分析

**正員、工修、トランスフィールド

(京都府京都市下京区中堂寺栗田町93番地、
TEL075-315-9186、E-mail hirotake@ird.co.jp)

***正員、工博、南山大学数理情報学部、教授

(愛知県瀬戸市せいれい町27、
TEL0561-89-2000、E-mail kawano@it.nanzan-u.ac.jp)

****学正員、東京大学大学院工学系研究科

(東京都文京区本郷7丁目3番1号、
E-mail junichi@nakano.gr.jp)

*****正員、工博、東京大学大学院工学系研究科、助教授

(東京都文京区本郷7丁目3番1号、
TEL03-5841-6235、E-mail hato@bin.t.u-tokyo.ac.jp)

2. 対象調査概要

本開発では、2005年11月21日～2006年1月22日の2ヶ月間、渋谷区域内を中心とした東京プローブパーソン調査(以下東京PP調査)のデータによって開発を行った。

以下に、東京PP調査概要・結果を記す。

(1) モニタ条件

モニタは、「関東近隣に居住」、「満15歳以上」、「月に1回以上渋谷に訪問する予定」の条件を満たしている50人を対象とした。男性25名、女性25名である。¹⁾

(2) システム特徴

東京PP調査では、従来のPP調査に比べ買い物の理由や広告看板のイメージといった語りに随時性のある情報取得が行える、PPブログシステムによる実装を行った。

このPPブログシステムの特徴として、従来のPP調査に加え、商品購入時、サービスを受けたとき、あるいは気になった広告を見た場所、気になった交通現象など、対象者の行動・活動動機・感情などのタイムリーなデータ(文章データ、写真など)を、携帯電話、WEBダイアリーによって残してもらい、対象者のより詳細な行動活動の把握が可能になる。

(3) 調査結果

東京PP調査では、都心回遊データ(以下トリップデータ)14456トリップ、随時性のある情報(以下エントリーデータ)13003エントリーが行われた。これは、PPブログシステムを実装によって、従来のPPシステムでのトリップデータの収集とともに、従来のPPシステムに比べエントリーデータの取得数の多さが特徴である。

図-1にエントリーデータ結果を示す。

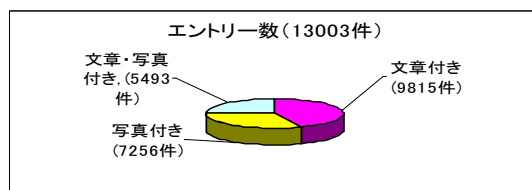


図-1 エントリーデータ

3. 時空間行動データ検索システム構造

昨今、インターネットの普及によって、インターネットやイントラネット上に存在するアクセス可能な膨大な文章から欲しい文章を探し出す上で、検索技術は不可欠の機能となっている。また、ブログなどが広く活用され始めると検索対象の特化、インターフェースの多様化が行われつつあり、様々な検索エンジンの開発、提供が行われている。これらの、検索技術の基本的な目的は、文章データの選択（あるいは絞込み）である。文章データの選択・絞込みは、優先的に分析対象にしたいデータや、情報精度の高いデータを得る方法として非常に効果的である。

同様に、詳細・大規模・高精度かつシームレスな時空間行動データの取得が可能なプローブパーソンシステムによって得られる時空間行動データの分析・評価を行う上で、対象データの絞込みは不可欠である。そこで、本開発では、プローブパーソンシステムによって取得可能な、特有な時空間行動データを、WEBダイアリーシステム、アンケート結果等による文章データを以下の手法により、並列化することによって時空間検索システムの構築を行った。

(1) 時空間行動データ・WEBダイアリーデータの並列化

本開発では、プローブパーソンシステムにより取得可能である質の異なるデータ郡である、時間、位置座標の連続による時空間行動データ、WEBダイアリーシステムで保管・保守され時間、位置座標によって時空間行動データに紐付けられている文章データをキーワード化することによって検索対象への変換を行った。

以下は、時空間行動データ、WEBダイアリーデータ（事前アンケートで得られる属性データを含む）に含まれる各データ郡のキーワード化手法である。

a) 位置座標データ

ポリゴンデータ作成ツールによって、任意のポリゴンデータを作成し、対象者のトリップOD、エントリー位置を内外判定によって検索キーワード化を行った。

b) 時間データ

時空間行動データが持つ時間データをトリップ、エントリー時間帯、曜日、天気などへの検索キーワード化を行った。

c) テキストデータ

エントリーに入力されたテキストデータを、全文検索エンジンによって検索キーワード化を行った。

d) 属性データ

調査前後アンケートによって取得された、各モニタの属性の文章データを各エントリーへの紐付けを行って検

索キーワード化を行った。

図-2に本時空間行動検索システムデータ構造を示す。

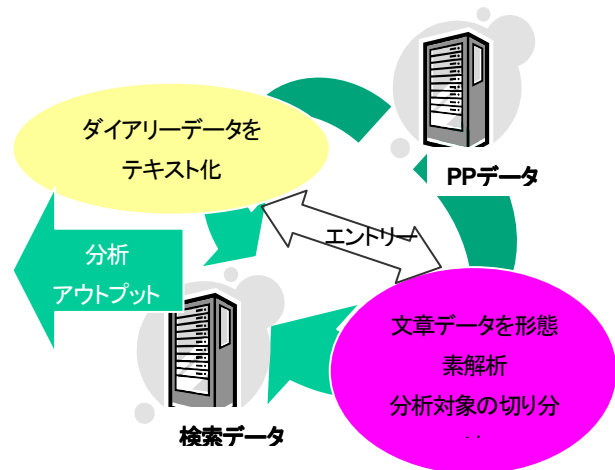


図-2 時空間行動検索システムデータ構造

(2) 時空間行動データ検索システム構築

図-3に本システムの時空間行動データ検索システム画面（イメージ）、表-1に各機能を示す。



図-3 システムの時空間行動データ検索システム画面

表-1 時空間行動検索システム機能一覧

サービス	機能
全文検索	任意の検索条件によるエントリーデータに紐付けられる時空間行動データの検索される
多次元分析	任意の検索条件で得られたデータの多次元分析を行いグラフ出力する(集計、他項目分析、時系列分析)
単語検索	任意の検索条件によるキーワードの出現数の計算を行う
バスケット分析	任意の検索条件によるキーワード間のアソシエーションコンフィデンスルールの算出・グラフ出力を行う
メンテナンス	シソーラス編集による検索データの保管・保守を行う

本検索システムは、キーワード、社会属性、時間帯、年齢、性別、天気、エリア、活動目的等によって、検索結果の絞込みを行い、検索されたエントリーデータは、詳細情報とともに、GIS上にマッピングして表示可能である。また、本システムには、検索結果によって得られた洗練されたデータの評価、解析機能、データの保管・

保守を行うメンテナンス機能の開発も行った。

4. 時空間行動検索システムの評価

(1) 検索ルールの絞り込み

渋谷区、新宿区における単語検索結果より出現率の高い単語を表-2に示す。新宿区域内のエントリーでは、「買う・購入」、「商品」、「買い物」といった購買に関係する単語の出現率が高い。対して、渋谷区域では、「買う・購入」は新宿区域に比べ低いが、「欲しい」といった、購買願望を表す単語の出現率が高くなっている。また、渋谷区域内では、「新宿」が出現率の高い単語として24件(1.5%)のエントリーに含まれているのに対して、新宿区域内での、「渋谷」を含むエントリーは出現率が低く12件(0.8%)にとどまっているため、新宿から渋谷への関係性よりも、渋谷から新宿の関係性が高いことが伺える。更に、「広告」、「おいしい」について、表-3に各属性率を示す。

表-3の結果より、「広告」の検索結果は、渋谷区域では、男性が74%を占めているのに対して、新宿区域では、女性の方が多く55%である。また、「おいしい」では、両区域で女性の割合が多いが、渋谷区域では、10代、学生の割合が多く、新宿区域では、40代、会社員・公務員の割合が高い結果が得られた。

このように、区域によって出現単語の特徴が表れている。これは、世代、性別などでの空間利用、行動形態の違いによるものと考察される。

表-2 区域選択による検索結果

渋谷区(エントリー:1555件)		新宿区(エントリー:1563件)	
単語	エントリー(件) 出現回数(回)	単語	エントリー(件) 出現回数(回)
広告	161(10.5%) 193	買った・楽しんだ	15
渋谷	103(6.6%) 113	学生	13
買う・購入	92(5.9%) 112	10歳代	11
良い	67(4.3%) 68	飲食店・喫茶店	10
食べる・飲む	39(2.5%) 43	広告看板	7
駅	38(2.4%) 41	会社員・公務員	6
クリスマス	33(2.1%) 37	40歳代	4
おいしい	29(1.9%) 31	金曜日	4
看板	28(1.8%) 31	新聞	24
目立つ	25(1.6%) 25	商品	23
変わる	25(1.6%) 25	買い物	23
新宿	24(1.5%) 24	本	19
友達	23(1.5%) 23	おもしろい	19
おもしろい	22(1.4%) 22	かわいい	19
プレゼント	20(1.3%) 21	欲しい	17
メール	20(1.3%) 21	凄いい	18
かわいい	18(1.2%) 19		
欲しい	17(1.1%) 17		

表3-1検索結果属性率 (広告)

区域	エントリー(件)	性別	年齢	職業
渋谷区	161(10.5%)	男性(74%)	40代(73%)	会社員・公務員(74%)
新宿区	163(10.4%)	女性(55%)	40代(80%)	会社員・公務員(94%)

表3-2検索結果属性率 (おいしい)

区域	エントリー(件)	性別	年齢	職業
渋谷区	29(1.9%)	女性(69%)	10代(38%)	学生(48%)
新宿区	35(2.2%)	女性(74%)	40代(69%)	会社員・公務員(80%)

(2) 絞り込み検索データ分析

表-4に「おいしい」についての渋谷区域内での性別毎のバスケット分析結果を行った数値の高い4件である。データ精度向上のため同義語として、「美味しい」、「うまい」を登録している。

表-4 バスケット分析 (おいしい、美味しい、うまい)

検索ルール	対象ワード	件数(件)	確信値(%)	サポート値(%)	リフト値(%)
女性【エントリー:20件】	買った・楽しんだ	15	75.0	2.73	2.23
	学生	13	60.0	2.18	1.05
	10歳代	11	55.0	2.00	1.15
	飲食店・喫茶店	10	50.0	1.82	5.38
男性【エントリー:9件】	広告看板	7	77.8	0.70	1.49
	会社員・公務員	6	66.7	0.60	0.78
	40歳代	4	44.4	0.40	0.77
	金曜日	4	44.4	0.40	2.88

表-4の結果より、渋谷区域内の女性においての「おいしい」に対しての対象ワードは「買った・楽しんだ」、「飲食店・喫茶店」など購買に関係の高い単語が見られ、感度が高い値を示している。対して、男性では、確信値はそれぞれ高い値を示しているが、「会社員・公務員」、「40歳代」においては、リフト値が低く感度が高いとは言いがたいが、「金曜日」などはリフト値も高く、渋谷区域内・男性の「おいしい」に対しての感度が高いと言える。

本章では、絞り込みによる各データの関係性の算出を行い、検索単語、エリアなどの検索ルールの変化による有効データの変化を確認した。

5. 終わりに

本開発では、プローブパーソン技術の発達により、時空間行動データ取得が定量的に取得できつつある中で、詳細・膨大で異なるデータ郡から、優先的に分析対象にしたいデータや、情報精度の高いデータを得る方法として非常に効果的である時空間行動検索システムの構築を行った。

本システムの利用によって、絞り込み、選択、共通化されたデータは、従来の属性によるセグメントではなく、十人十色な行動形態から十人十色な行動形態へ推移しつつある個人プロファイリングに有効な手段といえる。しかしながら、本システムはエントリーデータへの依存度が高いため、効果的にエントリーデータの取得方法の検討とともに、即時性のある効果的なデータ情報配信システムを構築していきたい。

参考文献

- 1) 三谷卓磨,二神雄典,高野精久,羽藤英二:プローブパーソン調査:展開の課題(2):オートエスノグラフィを基本にしたPPログシステムによる渋谷回遊行動分析,土木計画学研究・講演集,Vol.33, 2006 pp.18-32, 1994.