

F-039

行動情報に基づく広告価値マップ分析 Ad-Value Analysis Based on Behavioral Data

小池 亜弥† 白井 康之†
Aya Koike Yasuyuki Shirai

1. はじめに

GPSの普及により位置情報を収集する環境が整ってきた。位置情報は個人の行動パターンや嗜好を把握するために有用な情報であり、行動位置情報に基づくナビゲーションやリコメンデーション等、位置情報の活用に関する研究が盛んに行われている[1][2]。さらに、位置情報はその他のアクション情報(携帯電話利用履歴、おサイフケータイ等の購買履歴等)と組み合わせることで、行動目的や状況などを推測可能になり情報の価値が増す。

一方、モバイル広告の分野では、ユーザの携帯電話端末のGPS情報を基にその場所に関連する広告を配信する位置情報連動広告が行われている[3]。広告配信においては、広告に対してユーザが何らかのアクション(関連サイトの閲覧や商品の購買)を起こす反応率の向上が課題である。反応率向上のためには、よりユーザのシチュエーションを考慮した広告配信が重要になる。

本研究では約1,300人の25日間の携帯電話のGPSにより収集した位置情報と携帯電話の開閉履歴から、シチュエーションにおけるユーザの能動的携帯使用時間を求め「能動的に携帯電話を利用している程、情報に対する許容度があり、広告価値がある」という仮定の下、エリア毎の時間帯における広告価値を算出した。更に算出した広告価値の検証を行った。

なお、本研究は2009年度情報大航海プロジェクト「行動情報マイニングエンジンの開発」の一環で行われたものである。

2. 使用したデータ

使用したデータは2010年1月7日~1月31日の25日間に収集されたデータである。収集した情報は属性情報(アンケート)と行動ログデータである。

2.1 アンケートデータ

(1) 属性データ

属性データとして収集したデータ項目を表1に示す。

表1 属性データ項目

属性項目	選択肢
性別	0: 男性 1: 女性
年代	0: 10代, 1: 20代, 2: 30代, 3: 40代, 4: 50代, 5: 60歳以上
職業	0: 学生 1: 主婦 2: 会社員・自営業 3: 自由業・フリーター 4: その他
婚姻状況	0: 未婚 1: 既婚
家族構成	0: 単身 1: 夫婦のみ 2: 夫婦と子供 3: 3世帯家族

†(株)三菱総合研究所 Mitsubishi Research Institute, Inc.

	4:その他
所得	0: 200万円未満 1:200-400万未満 2:400-600万未満 3:600-1000万未満 4:1000万-1500万未満 5:1500万円以上 6:不明
居住地	自宅の郵便番号
勤務地	
常態地	勤務地またはその他常態地の郵便番号

(2) 携帯電話使用状況に関するアンケートデータ

携帯電話の利用状況に関して事前にアンケートで聴取した。聴取した項目は主に以下である。

- 1日のうちの自宅滞在時間、勤務先滞在時間、移動時間(平日、土日祝日ごと)
- 日常の中で携帯電話を能動的に(自分から)利用する割合
- 状況(日にち種類、状態、時間帯)ごとに携帯電話を能動的に利用する割合
- 携帯電話を利用する時間帯

2.2 行動ログデータ

(1) 位置情報ログ

位置情報は10分または30分間隔で収集したもので、分析したデータ件数は2,403,060件である。

(2) 携帯電話開閉ログ

携帯電話開閉ログは携帯電話端末の開閉の履歴であり、850,412件収集され、そのうち開ログと閉ログが1対1に対応し携帯電話使用時間を算出可能なデータのみを使用した。

3. 広告価値マップ分析方法

収集した行動履歴データを元に、駅単位で滞在する人の能動的携帯電話の使用時間を求める。携帯電話の利用には受動的利用(相手から電話やメールが来るから利用する場合)と能動的利用(自分から電話やメールを利用する場合)の2種類がある。広告価値マップ分析では「携帯電話を能動的に利用している場合に、情報許容度が高く、広告価値が高い」という前提仮定の下、駅ごとの広告価値の算出を行う。

広告価値マップ分析の算出フローは以下の通りである。

- (1) 携帯電話の利用履歴をもとに携帯電話の使用データに対して、状況を考慮し携帯電話を能動的に利用している確率を求める。
- (2) 駅ごとに能動的期待利用時間を集計する。

(3) 観測データが十分でない駅について、都市間類似度(外部データより算出)から補完し、推定広告価値を求める。

3.1 携帯電話能動的利用確率の算出

携帯電話を使用する状況を考慮し、使用状況において能動的に利用している確率 $P(\text{能}|状況)$ を求める。図1に考慮した状況変数を示す。状態とは、自宅滞在中、勤務先滞在中、移動中、その他のいずれかを表し、日にち種類は平日、土日祝日のいずれかを表す。

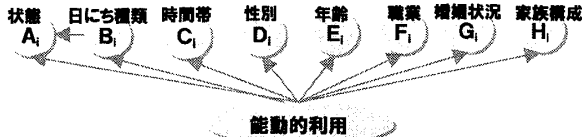


図1 考慮した状況変数

本分析では、「能動的利用の場合、状態と日にち種類は関連があるが、その他の状況変数間は独立」という仮定を置き、算出モデルの簡略化を行っている。

状態の算出ロジックを以下に示す。

<状態の算出ロジック>

- 判定1
自宅との距離が 300m 以下の場合、「自宅滞在中」と判定
- 判定2
勤務地との距離が 300m 以下の場合、「勤務地滞在中」と判定
- 判定3
移動速度が分速 100m 以上の場合、「移動中」と判定
- 判定4
上記判定1～3に当てはまらない場合、「その他」と判定

自宅との距離・勤務地との距離の算出には、位置情報と自宅・勤務先の緯度経度情報を元に、2地点の緯度経度から2地点間の距離を求めるヒュベニ式を用いて算出した。

携帯電話使用時の状況が与えられた場合の能動的利用の確率 $P(\text{能}|状況)$ を以下に定義する。添え字 k は個人を表す。

$$\begin{aligned}
 P(\text{能}|状況) &= P_k(\text{能} | A_i, B_j, C_l, D_m, E_n, F_o, G_p, H_q) \\
 &= \frac{P_k(\text{能}, A_i, B_j, C_l, D_m, E_n, F_o, G_p, H_q)}{P_k(A_i, B_j, C_l, D_m, E_n, F_o, G_p, H_q)} \\
 &= \frac{P_k(A_i, B_j, C_l, D_m, E_n, F_o, G_p, H_q | \text{能}) P(\text{能})}{P_k(A_i, B_j, C_l, D_m, E_n, F_o, G_p, H_q)} \\
 &\Rightarrow \frac{P_k(A_i, B_j | \text{能}) P_k(C_l | \text{能}) P(D_m | \text{能}) P(E_n | \text{能}) P(F_o | \text{能}) P(G_p | \text{能}) P(H_q | \text{能}) P(\text{能})}{P_k(A_i) P_k(B_j | A_i) P(C_l) P(D_m) P(E_n) P(F_o) P(G_p) P(H_q)}
 \end{aligned}$$

ここで、能動的な利用の場合に状態が A_i 、日にち種類が B_j である条件付確率 $P_k(A_i, B_j | \text{能})$ はベイズの定理により以下のように展開できる。

$$\begin{aligned}
 P_k(A_i, B_j | \text{能}) &= \frac{P_k(\text{能} | A_i, B_j) P_k(A_i, B_j)}{P_k(\text{能})} \\
 &= \frac{P_k(\text{能} | A_i, B_j) P_k(A_i | B_j) P_k(B_j)}{P_k(\text{能})}
 \end{aligned}$$

また、状態が A_i となる確率 $P_k(A_i)$ は以下で求める。

$$P_k(A_i) = \sum_{B_j} P_k(A_i | B_j) P_k(B_j)$$

式中の確率変数は、事前に行った携帯電話の利用状況に関するアンケートデータから算出した。

アンケートから算出した各確率変数の値は、ベイズ式を満たすため、以下の式を満たすように各確率変数の値を補正した。

$$\begin{aligned}
 P(\text{能}) &= \sum_{C_l} P(\text{能} | C_l) P(C_l) \\
 &= \sum_{A_i} \sum_{B_j} P(\text{能} | A_i, B_j) P(A_i | B_j) P(B_j) \\
 &= \sum_{D_m} P(\text{能} | D_m) P(D_m) \\
 &= \sum_{E_n} P(\text{能} | E_n) P(E_n) \\
 &= \sum_{F_o} P(\text{能} | F_o) P(F_o) \\
 &= \sum_{G_p} P(\text{能} | G_p) P(G_p) \\
 &= \sum_{H_q} P(\text{能} | H_q) P(H_q)
 \end{aligned}$$

$$\sum_{A_i} \sum_{B_j} P(A_i, B_j) = \sum_{A_i} \sum_{B_j} P(A_i | B_j) P(B_j) = 1$$

以上のプロセスにより、状況における携帯電話の能動的利用確率を求めることが可能である。

観測データにおける広告価値を能動的携帯利用時間と定義し以下で求める。

$$\begin{aligned}
 &\text{観測データにおける能動的携帯利用時間} \\
 &= \text{携帯利用時間} \times P(\text{能動的利用確率} | \text{状況})
 \end{aligned}$$

なお、こうした携帯電話の利用している状況(能動的/受動的)を推定する際には、ベイジアンネットワーク等のモデルも有効である。しかし、ベイジアンネットワークの場合、関係付ける全ての確率変数間の条件付確率を算出する為に、携帯電話の利用状況に関して事前にユーザに聴取する項目が膨大な数になってしまう。実サービスにおいては、ユーザに入力負荷ができるだけ少ない形での能動的形態利用時間の算出することが必要であるため、今回の研究ではベイジアンネットワークではなく、簡略化した確率推論モデルを用いた。

3.2 駅ごとの能動的期待利用時間の集計

観測データごとに求めた能動的利用時間を駅単位で集計したものを駅単位の広告価値として定義する。実験データへの依存性を軽減させるため、広告価値を求める条件の組み合わせにおいて、ユニークユーザ数が5人以上の駅(観測済み駅)のみ集計を行った。

広告価値は駅毎の単位時間あたりの平均能動的利用時間で表す。

$$\begin{aligned}
 &\text{条件} C \text{ での駅} A \text{ の広告価値} \\
 &= \frac{\text{駅} A \text{ の条件} C \text{ での能動的携帯利用時間合計}}{\text{駅} A \text{ の条件} C \text{ でのユニークユーザ数} \times \text{条件} C \text{ の時間帯の時間数}}
 \end{aligned}$$

ここで、条件Cの時間帯時間数とは、時間帯に含まれる時間数を表す。時間帯が朝であれば、5時～9時の4時間、求める時間帯が日中であれば9時～17時の8時間、求める時間帯が夜間であれば17時～24時の7時間、求める時間帯が深夜であれば0時～5時の5時間となる。条件別に相対比較をするため、時間帯の時間数で割り単位時間あたりの広告価値を求めた。

3.3 広告価値の補完

今回の実験では、モニタの人数・期間が限られていたため、データが十分に収集できていないエリアもある。データが十分に収集できなかった駅については、駅間類似度(外部データより算出)から補完し、推定広告価値を求める。

3.3.1 駅間類似度の算出

駅間類似度は、Mapion 電話帳¹から収集した駅単位のカテゴリ別施設数のデータ(表2に示したカテゴリを使用)を元に、駅間類似度を以下の方法で算出した。

表2 駅間類似度算出に使用した施設カテゴリ

No.	カテゴリ	No.	カテゴリ
1	グルメ	12	病院
2	デパート・百貨店	13	大学・短大
3	スーパー・コンビニ	14	学校
4	本屋・文房具	15	官公庁
5	電気屋・家電	16	大使館・領事館
6	カラオケ・ゲームセンター	17	ビジネス(写真・デザイン・装飾・録音)
7	美術館・博物館	18	ビジネス(建設・工事業)
8	公園・緑地	19	ビジネス(港湾・海運)
9	ホテル	20	ビジネス(倉庫業)
10	ラブホテル・モーター	21	ビジネス(水産・畜産業)
11	美容院・マッサージ	22	ビジネス(アパレル業服飾雑貨・趣味・民芸・工芸品)

駅Aのカテゴリ別施設数を列挙したベクトルを D_A 、駅Bのカテゴリ別施設数を列挙したベクトルを D_B とする。

$D_A = (A\text{駅のカテゴリ1の施設数}, A\text{駅のカテゴリ2の施設数}, \dots, A\text{駅のカテゴリ22の施設数})$

$D_B = (B\text{駅のカテゴリ1の施設数}, B\text{駅のカテゴリ2の施設数}, \dots, B\text{駅のカテゴリ22の施設数})$

駅Aと駅Bの類似度 $Similarity_{AB}$ を D_A と D_B のなす角と定義する。

$$Similarity_{AB} = \arccos \left(\frac{D_A \cdot D_B}{|D_A| |D_B|} \right)$$

3.3.2 広告価値の補完方法

データが十分に収集できなかった駅(未観測駅)について、データが十分に収集できた駅(観測済駅)の広告価値と駅間類似度からの推定広告価値の算出方法を示す。

広告価値分析では、同条件下でユニークユーザが5人以上の駅(観測済駅)については広告価値を求めた。ユニークユーザが4人以下の駅(未観測駅)については、データが十分でないため、推定広告価値を求める。

推定広告価値は、観測済駅の中で駅間類似度の上位10駅の広告価値の平均として求める。

4. 分析結果

4.1 属性による違い

4.1.1 職業による違い

求められた広告価値は属性の違いにより傾向の違いが見られた。中でも、職業による違いが大きく見られた。

図2～図4に職業ごとの平日日中の広告価値を示す。平日日中は、会社員自営業は広告価値が低い。一方で、主婦、フリーターの広告価値が高い。平日日中に就業しているか否かで時間的な余裕度に差があり、広告価値の傾向に違いが出ると解釈可能である。

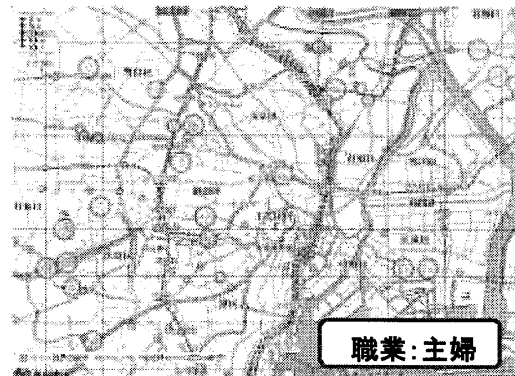


図2 平日日中の主婦の広告価値



図3 平日日中の会社員・自営業の広告価値



図4 平日日中の主婦の広告価値

¹ Mapion 電話帳 :

<http://www.mapion.co.jp/phonebook/>

4.1.2 性別・年代による違い

図5～図7に渋谷・表参道・外苑前の平日夜間の年代ごとの女性の広告価値を示した。平日の夜間、渋谷は10代女性の広告価値が高いが、表参道は20代女性の広告価値の方が高く、外苑前は30代女性の広告価値が高いと分かった。広告価値は誰が街に居るかではなく、街に滞在する人の中で誰の情報許容度が高いかを示している。今回の結果は年代毎にアフターファイブを楽しむ場所が異なっていることが背景にあると考えられる。

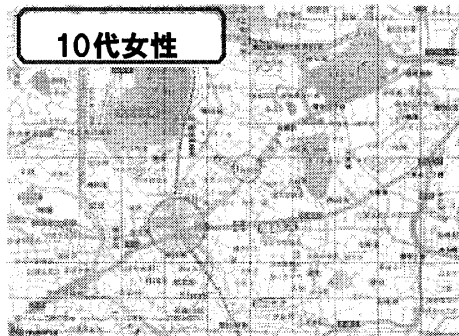


図5 平日夜間の10代女性の広告価値



図6 平日夜間の20代女性の広告価値



図7 平日夜間の30代女性の広告価値

5. 広告価値マップ分析の検証

ここでは広告価値推定手法の評価を行う。データをトレーニングデータとテストデータの2つに分割し、ト

レーニングデータで算出した広告価値によるテストデータの広告価値の予測能力を求めた。

5.1 検証方法

評価ではトレーニングデータでテストデータにおける広告価値の予測能力(汎化性)を見る。また、乗降客数の高い順に広告を出す場合と比較して、広告価値に従って広告を出す場合に想定される広告効率(効率性)を見る。トレーニングデータとして1,086人、テストデータとして265人のデータを用いる。

評価方法を以下に示す。

1. トレーニングデータで広告価値の高い順に駅を順序付け(トレーニングデータ広告価値順)。テストデータについて、以下の2パターンについて累積広告価値(累積比率)を比較したリフト図を描く。
 - ① トレーニングデータ広告価値順に広告価値を並べた場合
 - ② 乗降客数の高い駅順に広告価値を並べた場合
2. 広告の効率性として、乗降客数順に上位100駅を広告ターゲットとした場合の累積比率に対する広告価値順に上位100駅を広告ターゲットとした場合の累積比率の割合を広告価値リフト値として定義する。

なお、評価の対象とした駅は東京都内の駅のみである。広告価値順の累積比率が乗降客数順の累積比率を上回る場合、トレーニングデータでテストデータの広告価値を予測できる(汎化性が高い)、広告ターゲット選定の指標として乗降客数より広告価値を用いた方が広告効率が高い(効率性が高い)と評価する。

5.2 検証結果

図8に全てのデータを対象にした場合の広告価値のリフト図を示す。乗降客数準と比較して広告価値順が上回っていることより、広告価値の予測ができたと言える。広告価値リフト値より、広告価値を広告ターゲット選定の指標として用いた場合、乗降客数順に駅の規模を指標として広告ターゲットを選定した場合と比較して(上位100駅を対象とした場合)広告効率が1.665倍改善されると解釈できる。更に、属性を限定した場合、より予測能力が高くなった。

図9は性別を男性に限定した場合のリフト図を示すが、リフト値は6.669倍と全体平均を大きく上回った。その他曜日や時間帯等を限定した場合もリフト値は全体平均を大きく上回り、条件を限定すると広告価値の予測能力が高まると考えられる。

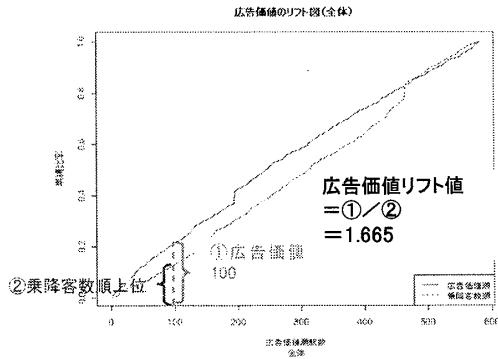


図8 広告価値(全体)のリフト図

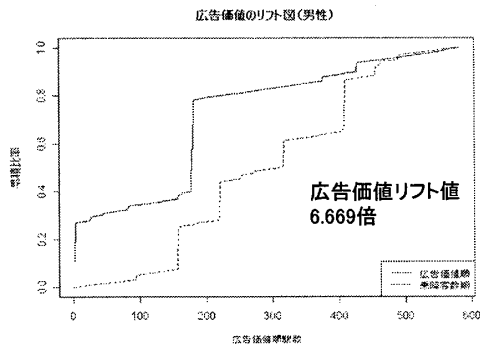


図9 広告価値(男性)のリフト図

更に、位置情報・携帯電話端末の開閉情報のみならず、メールの閲覧や地図情報サービスの利用等のアクション情報も収集可能となれば、ユーザの嗜好や行動目的等も把握することができ、より効果的な分析が可能になると考えられる。今後よりリッチなアクション情報との融合が期待される。

謝辞

本研究は 2009 年度情報大航海プロジェクト「行動情報マイニングエンジンの開発」の一環で行われたものである。

参考文献

- [1] 篠田 裕之, 竹内 亭, 寺西 裕一, 春本 要, “行動履歴に基づく協調フィルタリングによる行動ナビゲーション手法”, 情報処理学会研究報告.GN,[グループウェアとネットワークサービス], 2007(91),87-92, (2007).
- [2] 佐藤 一夫, 山本 真人, 小林 功, 佐治 信之, 田中 克明, “行動履歴に基づく情報推薦基盤と推薦エンジンの開発”, 電子情報通信学会技術研究報告. AI, 人工知能と知能処理, No.108, 33-38 (2008)
- [3] シリウステクノロジーズ, 位置連動広告「Ad Local」
<http://www.cirius.co.jp/adlocal/>

6. まとめ

本研究では携帯電話端末の位置情報履歴と開閉履歴、携帯電話使用状況から携帯電話能動的利用時間（自分から積極的に携帯電話を使用している時間）を推定する予測モデルを構築した。トレーニングデータとテストデータを用いた予測能力検証においては、殆どの属性において提案モデルの予測能力を検証できた。

提案モデルの応用例としては、携帯電話向け広告への利活用が考えられる。広告価値の高い層を広告ターゲットとすることで、時間と場所に応じた携帯電話向け広告の配信最適化ができ、広告効率の向上が期待できる。

また、マーケティングへの活用も想定される。例えば新商品のサンプリングの際のターゲットエリアの選定が可能である。一般的には乗降客数等の街の規模等でターゲットエリアを選定していると想定されるが、広告価値を用いることで時間帯や属性ごとにより反応率が高いエリアを効率的に選定することが可能である。

このように、提案モデルの分析結果は広告・マーケティング等多岐に渡り活用が期待できる。

今後の研究の発展性としては、今回前提においた「携帯電話を能動的に利用する時間＝広告価値が高い」という仮説について、検証する必要がある。今後、アンケートや実サービスの中でシチュエーションごとに感情等をユーザに入力してもらうことが可能であれば、実際のシチュエーションにおける携帯電話の能動的利用と広告価値との適合度を検証していきたい。

また、今回の実証データでは、人数・期間の不足から、多人数が集まる繁華街以外では、あまり有効なセグメントが抽出できなかったとも考えられる。今後、更に大規模なデータ収集を行うことが課題である。