

需要・購買行動を監視する予実監視ダッシュボード

宇山 政志[†] 渡辺 理[†] 織田 充[†] 松本 安英[†] 神田 陽治[‡]

[†]富士通研究所 [‡]富士通

〒674-8555 明石市大久保町西脇 64 〒211-8588 川崎市中原区上小田中 4-1-1

あらまし Webサイト上で、キャンペーン等のイベントを実施する場合に、来訪者数および購買意思の変化を予測する行動モデリング技術、および、予測と実績の差異を、日次単位・10分単位で監視し、差異原因を分類し、監視者（運用担当者およびマーケティング担当者）の施策決定を支援する予実監視ダッシュボードを提案する。

キーワード 需要予測、予実監視、知識共有、データセンター、行動ターゲティング、サービスサイエンス

A Dashboard-style Demand Monitoring Tool for Web Sites

Masashi Uyama Satoru Watanabe Mitsuru Oda Yasuhide Matsumoto Youji Kohda

FUJITSU LABORATORIES LTD. 64, Nisiwaki, Ohkuho-cho, Akashi 674-8555, Japan

Abstract This paper proposes a set of user behavior modeling methods, a demand forecasting method, and a dashboard-style demand monitoring tool for Web sites. The forecasting method predicts how campaign events affect both web access variation and purchase behavior variation. The dashboard visualizes the differences between demand forecasts and actual demand per day or per 10 minutes and assists the marketing decision making.

Keyword Demand Forecasting, Demand Monitoring, Knowledge Sharing, Datacenter, Behavioral Targeting, Services Science.

1. まえがき

現在、オンラインバンキング、オンライン書店など多数のサービスがインターネット上で提供されている。サービスの利用拡大のために、割引キャンペーン等の突発的なアクセス増加を伴うマーケティング施策がしばしば企画される。キャンペーンを企画する上では、(1)どのような種類のインセンティブを、(2)どれだけの量提供し、(3)どれだけの広告を出稿し、(4)どれだけの期間続けるかを、投資対効果に見合うように計画する必要がある。利用者層の関心や嗜好は絶えず変化しており、過去の事例と全く同じ条件でキャンペーンを行ったとしても、以前と同じ効果は期待できない。このため、キャンペーン期間中でもお客様の反応を見ながら動的に施策を変更し、当初予定した目標に近づけるように微修正していくことが必要になる。

予測値と実績値を単純に比較して、乖離を示すことができたとしても、それだけでは、乖離原因を推定することはできず、したがって有効な対策の示唆を行う

ことはできない。このため、キャンペーン計画者の勘と経験にもとづき、施策を変更するしか手段がなかった。我々は、マーケティング施策の効果を監視するためには、マーケティング担当者と運用担当者との間で、知識の共有が必要と考えている。キャンペーンを企画するマーケティング担当者は、Webサイト構造の意味や、キャンペーンの狙い・意義について知識を有する。一方、データセンターの運用担当者は、利用者行動をリアルタイムに、複数サービスに渡り把握できる利点がある。Webサイトの運用において、マーケティング担当者の知識、運用担当者の知識を相互に活用し、マーケティング施策およびITリソースの最適化を実現することが、本研究の目的である。

2. 従来のWebサイト監視

現状でもWebサイトを監視する様々なツールが存在する。しかし、マーケティング担当者向けの監視と運用管理者向けの監視が分離し、両者の知識が共有されていない点に問題があると考えられる。

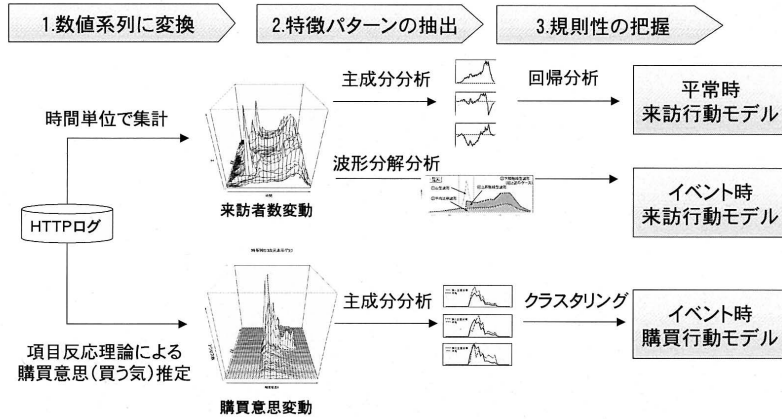


図 1 行動モデリングのステップ

市販のログ解析ツールを用いれば、広告媒体ごとのコンバージョン率や、ページ遷移の頻度などを日次単位で集計・表示できる。しかし、表示の意味を解釈し、何が起きているかを判断するには、Webサイトに関する詳細な知識が必要である。例えば、「今日は、いつもより、AA/BB.html -> CC/DD.html への遷移が3倍増加している」ことがわかったとしても、サイトの内容に詳しくない第三者はその意味を理解できない。一方、運用者向けの監視ツールは、CPU負荷等の監視やアラートが中心で、利用者のサイト上での行動までは踏み込んでいない。

3. 行動モデリング技術

我々の行動モデリング技術では、サイト内容に詳しくない運用担当者でもサイトのマーケティング施策に対する監視／アラートが可能になるように、1) 利用者の複雑な行動ログ (HTTPログ) を、数値系列に一旦変換し、2) 主成分分析等の統計手法を用いて特徴的なパターンを抽出し、3) 回帰分析等を用いて各パターンが生起する条件の規則性を把握する (図1)。

3.1. 平常時来訪行動のモデル化

N日分のHTTPログから利用者セッションの開始数を10分単位で集計することで、来訪者数の変動を、144点×N日の数値マトリクスで表すことができる。この数値マトリクスに対して、統計手法を用いたモデリングを行い、イベントのない平常時の日内変動の傾向をモデル化することができる。

文献[8]で提案した手法では、144点×N日の数値

マトリクスから、異常日を除いた後、主成分分析を適用することで、利用者の日内行動を特徴的な行動パターンに分解している。抽出した行動パターンの日次推移に対して、回帰分析を適用し、各行動パターンの増減に影響を与える日種因子を選定する (図2)。日種因子とは、「土曜日」「祝日」「4月」「お盆」などに該当する場合1、そうでない場合0となる離散変数である。日種因子に対する利用者の行動パターンの増減をモデル化することで、中長期の平常時の需要予測に利用できる。

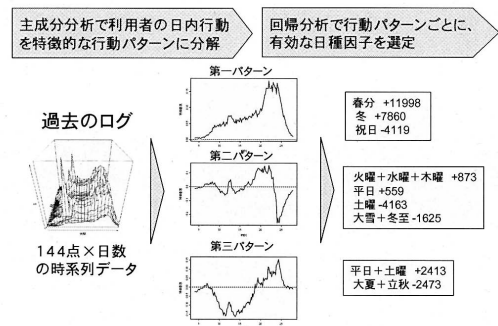


図 2 平常時の来訪行動のモデル化

以上のモデル化はプログラムで自動的に行われる。さらに、マーケティング担当者は、該当サイトの利用者層に関する知識を動員して、主成分の意味を解釈することができる。例えば、各主成分に対して「アフターファイブ型行動」「夜更かし型行動」といったラベル付けを行うことができる。需要を予測する際に、「夜更

かし型行動が2倍」になった想定外の需要データを用意して、運用担当者に伝達しておくこともできる。運用者側では、実績値が通常の予測を超えて、想定外需要データに近くなれば、「アフターファイブ型行動が想定外に増えている」といった形でアラートをあげることができる。

3.2. イベント時来訪行動のモデル化

文献[1][2]では、キャンペーン等のイベントを実施したときのサイト来訪者数の時系列を、1)山型の波形、2)平均に連動した波形、3)階段状に下降する波形、4)階段状に上昇する波形に分解して、特徴を抽出している。各波形の特徴を、波形毎の人数比など少数のパラメータでモデル化し、当日実施していたイベントの情報と紐付ける(図3)。

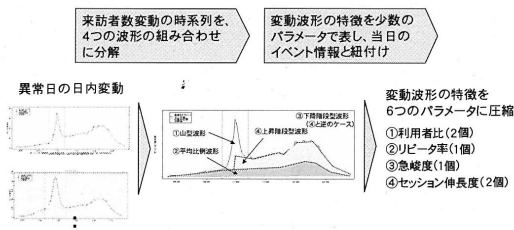


図3 イベント時の来訪行動のモデル化

運用者側では、多数のサービスに関して、イベントに対する来訪者の反応のパターンをモデルして蓄積しておくことができる。多くの事例を蓄積することで、新しいキャンペーンを行う場合でも、過去に実施された様々なキャンペーンの蓄積事例から、類似の事例を検索し、来訪者数の遷移を予測することが可能になる。

3.3. イベント時購買行動のモデル化

Webログ分析の分野で、利用者のページ参照と、購買行動の関連を表す指標として、コンバージョン率がある。コンバージョン率は特定のページ(もしくは最初に参照したページ、あるいはサイトに誘導したバナー広告)を参照した来訪者が、最終的にサイトの目的とするページ(コンバージョンページとも呼ばれる)に至る確率を示す。しかし、コンバージョン率では、来訪者がキャンペーンに期待はずれで帰ったか、関心はあるが最終決断を見送ったのかの違いはわからない。マーケティング施策を考える上では、購買には至らなくても購買意思の高い見込み客がどの程度来訪しているかを知る必要がある。

Webサイト上のページ参照履歴を追跡すれば、その利用者の購買意思をある程度推測可能である。しかし、そのためには、Webサイトの個々のページの役割を知っておく必要があり、サイト内容に関する知識を持たない第三者にとっては困難である。

本研究では、Webサイト上での利用者の行動を、購買意思の度合いを表すスコアで数値化する。数値化に用いるのは、テスト問題の作成等に利用される項目反応理論(Item Response Theory: IRT)[5]である。Webサイトの行動分析に、項目反応理論を用いることは、既に橋間ら[6]により報告されている。項目反応理論では、受験者の能力が高いほど問題の正答率が高いという前提に基づき、各問題の難易度をS字曲線(ロジスティック関数等)でモデル化し、S字曲線モデルと各受験者の正答パターンから、受験者の能力および問題の難易度を同時に推定する。

我々のモデリング手法では、サイトの目的ページのURLを与えると、テスト問題に相当するURL集合を過去ログから自動抽出する。利用者が各URLを「参照した・しない」の情報をテスト問題の正答・誤答とみなし、各セッションのURL参照パターンを項目反応理論に基づいて分析し、潜在的な購買意思の高さ θ を推定する(図4)。数値化しておくことで、サイト内容に詳しくない第三者でも、購買意思の予実比較/アラートを行うことができる。

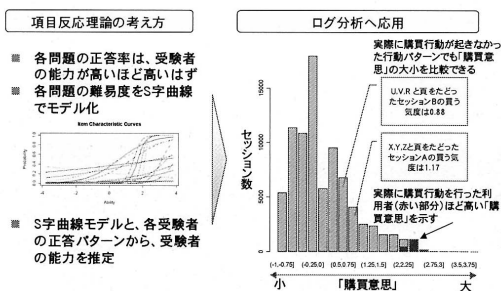


図4 項目反応理論を用いた購買意思推定

次に、キャンペーン等のイベントが、利用者層の購買意思分布に与える影響をモデル化する。インセンティブの種類が異なっても、イベントが購買意思の変化に与える影響は、一定のパターンに分類できると考えている。このパターンをイベント効果型と呼ぶことにする。

N日分のログがあれば、利用者層の購買意思 θ の分

布N日分の数値マトリクスが生成できる。数値マトリクスに主成分分析を適用し、少数の特徴的な変動パターンを取り出す。次に、キャンペーン等のイベントを実施した際の効果を、k-means等のクラスタリング手法を用いて、変動パターンの一定の組み合わせ（イベント効果型）に類型化することで、イベントが起きたときの購買意思の変化をモデル化する（図5）。

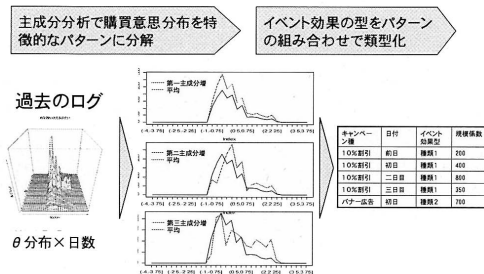


図5 イベント時の購買行動のモデル化

以上のモデル化はプログラムで自動的に行われる。キャンペーン内容に関する知識を有するマーケティング担当者は、分析結果を解釈し、各主成分およびイベント効果型に対してラベル付けを行うこともできる。運用者側では予実の差異に対して、このラベルを用いて、アラートをあげることもできる。

4. 予実監視ダッシュボード

過去の事例から、平常時の来訪行動、イベント時の来訪行動、イベント時の購買行動がモデル化できていれば、将来類似のイベントを行うときの来訪者数を予測し、利用者の購買意思の分布を予測することができる。予測した来訪者数・購買意思の分布と、実績値とを日次単位・10分単位で、比較・表示を行うのが、図6に示す5つの画面からなる予実監視ダッシュボードである。

- ・第1の画面は、来訪者数の「日内変動」の予測と、実績とをリアルタイムに10分ごとに監視する画面である。予定したイベントに対して、想定内需要と想定外需要の2種類のデータをあらかじめ用意し、運用中の需要データと比較して、想定外の変動があったときにアラートをあげる[3]。
- ・第2の画面は、来訪者数および購買意思分布に関して、前日の予測と実績の比較、および、当日・翌日分の短期予測データを表示する画面である。
- ・第3の画面は、来訪者数の約150日分の中期予測と、予測に従ったサーバ配備計画、および過去の予測と実績の精度を表示する画面である。
- ・第4の画面は、前日の購買意思分布実績値を、イベント時購買行動モデルの主成分ごとに分解し、異常原因を探るための画面である。
- ・第5の画面は予実を比較して、自動的にアラートをあげるコンソール画面である。

本ダッシュボードを用いれば、サイト内容に詳しくない運用担当者でも、第1～第3の画面を監視し、キャ

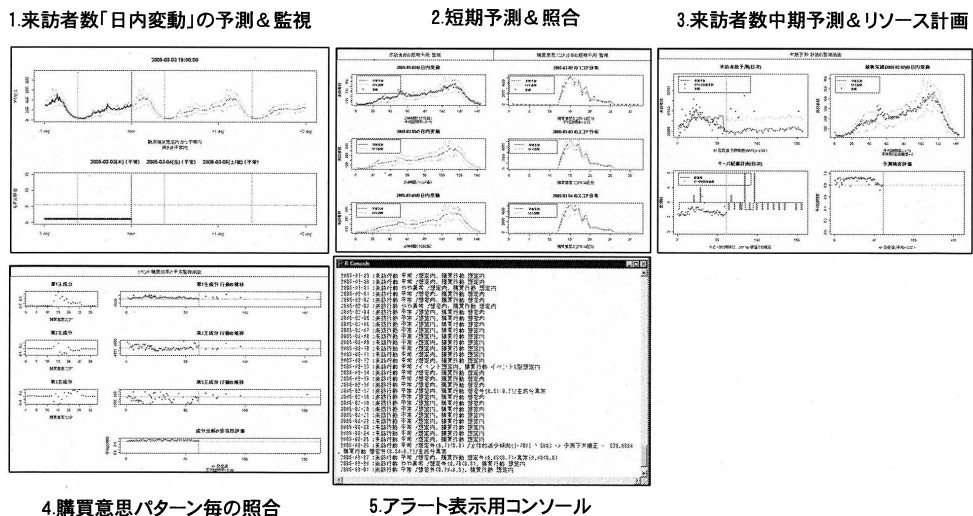


図6 予実監視ダッシュボードの5つの画面

ンペーンにより来訪者数や購買意思が予定通り変化しているかどうかアラートをあげることができる。また、予実の乖離があったときには、第4第5の画面を見ることで、乖離の原因を知ることができる。

例えば、購買意思の分布に、予測と実績の乖離がある場合に、乖離の原因が規模の増減か、特定の主成分の増減にあるかを判断できる。規模の修正のためには、広告規模の拡大により、キャンペーン自体の規模を補正する施策が適当である、単一主成分の増減の場合には、該当主成分を特に強化する施策を過去の事例から探索し、追加実施するといった判断が可能になる。

5. 結び～知識の共有に向けて～

行動モデリング技術と、予実監視ダッシュボードを用いたWebサイト監視での、知識共有のアプローチを図7にまとめる。運用者はリアルタイムで利用者の行動を把握できるが、Webサイトやキャンペーンに関する知識はない。マーケティング担当者側で持っているサイト構造、サイト利用者、イベント実施等の知識は、行動モデリングの過程で、購買意思指標・主成分のラベル・規則性の紐付けなどの形で、モデルの中に埋め込まれ、需要予測に反映される。予実監視ダッシュボードでは、運用者は予実の数値のみを比較すればよい。予実の乖離があった場合は、主成分の予実差異にまでさらに分解・ドリルダウンして比較することで、原因の推定や施策提案まで行うことができる。また、運用者側では、複数のサイトにおけるイベント実施の事例を分析・蓄積することで、マーケティング担当者が実施する類似キャンペーンの予測を支援することが可能になる。

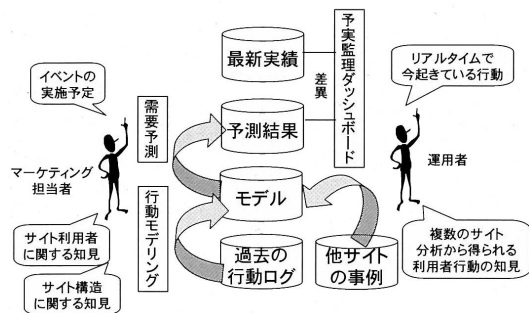


図7 知識共有に向けてのアプローチ

データセンターには、様々なサービスの運用ログが集積される。運用担当者は、利用者行動をリアルタイムに、複数サービスに渡り把握できる利点を持つ。このため、様々なサービスの運用ログを活かし、新たな

サービス価値を創出しようという考えがある[4]。また、データセンターの運用ログを収集する仕組みをソフト的なセンサと捉えて、このようなソフトセンサと、カメラやRFIDのようなハードセンサとを組み合わせ、業務の最適化を図るという考え方もある[3]。

本稿が主張するのは、上述のようなサービス価値の創出や、業務最適化といった目的を実現するためには、関係する人々の知識を共有する枠組みが不可欠である、ということである。運用ログやセンシングデータの中には様々な有用な情報が存在する。しかし、これらの情報は、知識を持つ人が解釈することで、初めて利用可能になると考えるからである。

本稿で提示したダッシュボードは、複数サイトの分析から得られたモデルを合成した仮想的なサイトでシミュレーションを行ったものであり、実際のサイトでの予実管理に適用したものではない。今後、実際のサイトのログに適用し、ダッシュボードを用いて、ITリソースの最適化だけでなく、マーケティング施策の最適化も同時に可能であることを検証していく必要がある。

文 献

- [1] 宇山政志、松本安英、矢崎昌朋、神田陽治、勝山恒男、「利用者行動モデルに基づくイベント性需要変動予測」、電子情報通信学会第5回Webインテリジェンスとインタラクション研究会、2006.3, pp.77-82.
- [2] 宇山政志、松本安英、渡辺理、矢崎昌朋、神田陽治、「Webサイト利用ログからの需要変動推定」、電子情報通信学会第8回Webインテリジェンスとインタラクション研究会、2007.3, pp.91-96.
- [3] 勝山恒男、神田陽治、「現場業務を実行・最適化するサービス基盤技術」、F U J I T S U, Vol.58, No.3, 2007.5, pp.234-239.
- [4] 神田陽治、勝山恒男、丸山文宏、「サービス価値の創出へ向けて」、人工知能学会誌 2007年11月号 (Vol.12, No.6), pp.763-770.
- [5] 豊田秀樹、「項目反応理論(入門編)」, 朝倉書店、2002
- [6] 橋間智博、「IRTによる情報提供型Webサイトのアクセス分析」、FIT2006(第5回情報科学技術フォーラム), pp.423-425、情報処理学会 (IPSJ) と電子情報通信学会・システムソサイエティ (ISS) 及びヒューマンコミュニケーショングループ (HCG) 合同主催、2006.9.5.
- [7] 松本安英、宇山政志、矢崎昌朋、神田陽治、「Webサーバへの想定外需要検出方式の実装と評価」、情報処理学会第16回システム評価研究会、2006.3, pp.1-6.
- [8] 渡辺理、神田陽治、矢崎昌朋、宇山政志、織田充、松本安英、「Eコマースサイトの利用者行動の解明：日種因子による主成分回帰分析」、情報処理学会マルチメディア、分散、協調とモバイル (DICOM2006)シンポジウム、2006.7, 3C5.