

インタラクティブシステムにおけるユーザー特性の分析方法

石田 奈緒子 犬束 敏信 篠田 晃

NTT ソフトウェア研究所

NTTではマルチメディア通信の共同利用実験の一つとして、CATV映像伝送等利用実験を行っており、実際に一般家庭でインタラクティブテレビを利用することが可能である。今日では一家庭を対象にしたマルチメディアの導入例は少なく、今後のマルチメディア技術の発展のためには家庭生活への影響を調べるが大変重要である。今回の実験では、ユーザが行った操作履歴からユーザの振る舞いを知ることができるため、それを分析することにより家庭生活への影響度を知ることができる。我々はユーザの振る舞いを解析するツールを使用し分析を試みている。本稿では、ツールを用いた分析方法、適用例を報告する。

1. はじめに

インタラクティブ・サービスでは、ユーザはサービスを受けるだけでなく、その意思によりシステム（サービス）に働きかけることができる。そのため、その働きかけの状況を把握することにより、ユーザの振る舞いを知ることができるのだが、これにはユーザの癖や嗜好といった個性が含まれる。そこで、このユーザの振る舞いを解析すれば、ユーザの利用特性を分析でき、これをサービスにフィードバックすることで、サービスを改善することも可能となる。

現在、一般にマルチメディアという場合、パソコンを中心とした個人向けのものが中心であり、一家庭を対象にしたマルチメディア導入例は少ないため、今後のマルチメディア技術の発展のためには、マルチメディアの家庭生活に対する影響を調べることは大変重要である。

以上より、今後より良いインタラクティブ・サービスを展開するためには、ユーザの振る舞いを解析しサービスにフィードバックする機能を備えることが必要となる。

本実験では、ユーザの特性を調査すべく機能を備えているので、その機能を実現したツールを利用した分析方法と適用例について述べる。

2. 分析システム構成

今回の実験で使用しているインタラクティブ・システムは、クライアント／サーバー構成である。サーバーはセンターに設置し、デジタル映像データの蓄積、ユーザ情報や課金情報の管理等を行なっている。クライアントは、ユーザ宅に設置したセットトップボックス(Set Top Box:STB)であり、ユーザはSTBに対してチャンネルの選択、映画等の映像コンテンツの購入操作を行い、サーバーのあるセンターと通信することでサービスを享受できる。

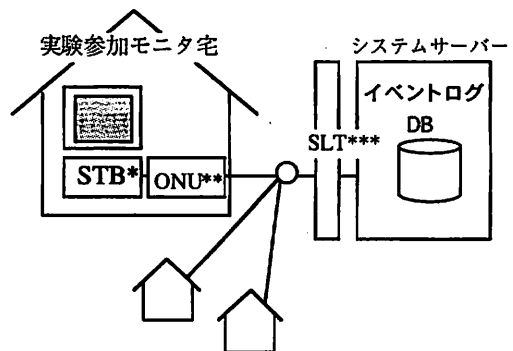


図1：実験システム概要図

*光網終端装置、**セットトップボックス、***加入者線端局装置

ユーザの振る舞いを分析するには、STBでユーザの操作を情報として記録、蓄積する必要がある。本実験システムでは、ユーザによるSTB操作の一つ一つをイベントの発生ととらえ、発生の度にシステムサーバーにそのイベント情報を送信し、ログとして蓄積されるしくみを採用している。ここで蓄積されるログはイベントログと呼ばれ、STBのIDや契約者のID、日時、サービス（チャンネル）名、イベント名等の情報が含まれている。そのため、ユーザの振る舞いの情報は、イベントログを使用することとした。

文献 [1] では「カスタマの振るまいを多面的に分析し、結果をサービスへフィードバックする機能を Preference Analyzer と呼ぶ」としているが、本実験においては、収集したイベントログを使用し Preference Analysis Model [1] を用いたデータ解析を行う機能を Preference Analyzer（以下PA）と呼ぶことにする。PAでは、イベントログが蓄積されたイベントログ・データベースを元に、データを解析し、結果をテーブルとして出力する（図2）。以下、解析結果を得るまでの各処理について述べる。

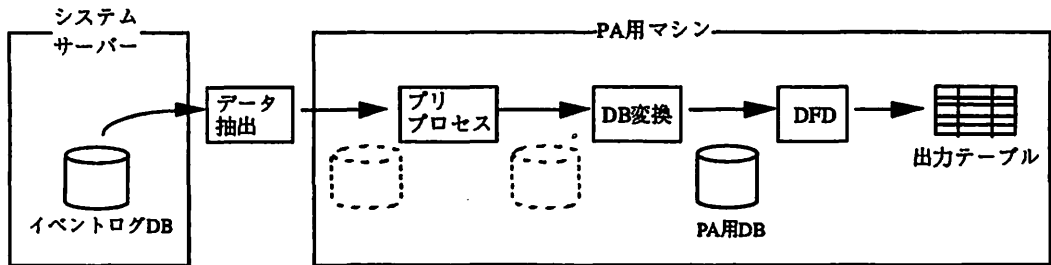


図2：PA関連データベース構成

(1)プリプロセス

まず、イベントログDBに蓄積したデータを、ワーク用DBに必要な量を取り出し転送する。このときの量は、解析単位である。例えば、週間分、月間分等である。

次にワーク用DBのデータを解析する前処理として、プリプロセッサにかける。これは、ユーザの操作情報であるイベント情報は、イベント前後で相関があり、ある条件に従いワーク用DBに抽出したデータはデータの開始と最後で相関がとぎれている等、不完全な場合があるため、これを補正するためにプリプロセッサを使用する。

(2)データベース変換

PA分析用のデータベースを構築する。

プリプロセスを実行したところまでのデータベース内のログは、イベントが発生した瞬間に出されるログ（瞬間イベント）であり、ある処理時間（停滞時間）のイベントの状態を示すものではない。データベース変換とは、瞬間イベントを、その状態に移移した時刻、その状態から他の状態に移移した時刻、その状態に停滞していた時間を付加する等、サービスの状態遷移と対応づけたログ（状態イベント）に加工し、PA内のイベントログ・データベースに格納するプロセスである。この状態イベントと、一部加工された瞬間イベントは別々のテーブルに格納される。

(3)解析処理

Preference Analysis Model [1] を用いてデータ解析を行う。

解析ツールは Win32 環境上で動作し、GUI を使用し、データフロー図 (Data Flow Diagram:DFD)により解析処理を記述し実行さ

せるインタープリターであり、解析結果としてデータテーブルを出力する。今回の実験で現在、解析により得ている情報は 33 種 (表 1) であり、それぞれの情報に対し DFD を作成して解析行う。最後に出力されたデータテーブルを図表化する。この作業では表計算ソフトを使用する。

表 1：PA を使った分析の実際

チャンネル	テーマ	内容
全体	曜日別アクセス数/購買数	当月のインタラクティブ ch/有料コンテンツへの総アクセス数を、曜日別に一日あたりに平均したもの
全体/カラオケ/さわる CM	時間帯別アクセス数/購買数	当月のインタラクティブ ch/有料コンテンツへの総アクセス数を、1日の時間帯別に平均したもの
全体	ch 別アクセス数月別推移	月別の各インタラクティブ ch へのアクセス数推移
全体	世帯年収別平均月額料金	同じ年収 (世帯年収) の世帯グループ毎に、当月のインタラクティブ ch 月額料金を平均したもの
全体	平均月額料金推移	インタラクティブ ch 月額料金の推移
全体	職業別平均月間アクセス数	同じ職業のユーザのグループ毎に、当月のインタラクティブ ch アクセス数を平均したもの
全体	CATV/ITV/EPG の一日の視聴時間	CATV/インタラクティブテレビ/番組ガイドの一日の視聴時間の長さ毎に、世帯をグループングしたもの
全体	パソコン・パソコン通信との相関関係	当月のインタラクティブ ch の利用について「パソコンの利用経験がある人/ない人」「パソコン通信を利用している人/利用していない人」別の、平均月間アクセス数
全体	ch 別アクセス率	インタラクティブ ch 総アクセス数における、各サービスのアクセス (30 秒以上) 数の割合
全体/各サービス	年代別アクセス率	年代別のユーザ数にばらつきがあるので、各年代の総ユーザ数を 100% とし、そのうち当月にインタラクティブ ch を利用したユーザの割合
全体/各サービス	月別アクセス数/購買数推移	月別のインタラクティブ ch/有料コンテンツへのアクセス数の推移
全体	利用世帯率	全世帯を 100% とし、当月にインタラクティブ ch を利用した世帯の割合
全体/各サービス	一日当たりの利用時間別世帯比率	当月の一日あたりのインタラクティブ ch 平均利用時間別の世帯数比率
各サービス	月額料金別世帯数比率	当月の月額料金別の世帯数の比率
各サービス	トップ 10	当月に購買 (視聴) されたコンテンツのうち、上位 10 作品のタイトルと総アクセス数
MOD	レンタルビデオとの相関	レンタルビデオのレギュラー/ライトユーザ/非ユーザのグループ毎の月間平均購買数

チャンネル	テーマ	内容
		*レギュラー：週1本以上、ライトレギュラー：2週間に1本以下、非レギュラー：ほとんど借りない
SOD/COD	カテゴリ別アクセス率	当月のコンテンツ購買数（視聴数）の、カテゴリ別の割合
EPG	利用時間率	当月の EPGch へのアクセス数の、利用時間の長さ別の割合
EPG	利用率	当月の STB 経由の CATVch アクセスの内、EPG を利用したものの割合
EPG/さわるCM/大人の書斎	曜日別アクセス数	当月の EPG へのアクセス数の曜日別一日平均
EPG	テレビ番組情報誌と EPG 利用の 相関関係	各種テレビ番組情報誌を利用している世帯/利用していない世帯の、EPG への平均月間アクセス数
ITV ナビゲータ	各インタラクティブ ch への移行数	当月に各インタラクティブ ch 紹介画面から直接その ch へと移ったアクセス数
カラオケ	他カラオケサービスとの相関	他カラオケサービス（お茶の間カサネ＆お茶等）のレギュラー/ライトレギュラー/非レギュラーのグループ毎の月間一人あたりの購買アクセス数
カラオケ	一回当たりの利用時間	当月の利用時間の長さ別の割合
カラオケ	利用形態	カラオケ ch で利用者に回答をお願いしているアンケート結果（利用形態：一人で/家族だんなに/友達と/その他）
ショッピング	通信販売の利用との相関	通信販売を利用している/利用していないユーザの、ショッピングへの平均アクセス数
コミュニティ	地域別アクセス数	当該商店街を選択したアクセスの、町名による分類
TBSインタラクティブ	各メニューのアクセス数	メニュー画面での、各メニューへのアクセス構成率
TBSインタラクティブ	各メニューの移行率	各メニューでの、カテゴリ別アクセス構成比
さわるCM/大人の書斎	他 ch への移行率	さわるCMの次にアクセスしたchの構成比
さわるCM	コンテンツアクセス数	各コンテンツへのアクセス数の構成比
さわるCM/大人の書斎	コンテンツ利用時間	各コンテンツ/書籍への1アクセス当たりの平均利用時間
大人の書斎	タイトルアクセス順位	各書籍へのアクセス数を多い順に並べたもの

3. PA による分析作業

この章では「チャンネル別アクセス率」を求める場合を例に、分析方法を示す。

PA において重要なのは DFD の作成である。

DFD はノードとアークから構成される。解析ツールでは3種類のノード（表2）、さらに演算用に11種類のノードが用意されている（表3）。

表2：PA の記述法

記号	機能
アーク	ノードからノードへのデータの流れを示す
ノード	
データベースノード	データベースからテーブルを取り出す機能をもつ
出力ノード	アークから流れてきたテーブルをファイルに出力する
演算ノード	アークから流れてきたテーブルを加工する

表 3：演算ノードの内容

ノード名	内容
選択ノード	指定した条件を満たすレコードを抽出する
カテゴリノード	指定した条件を満たすレコードにタグを付与する
フォーカスノード	指定したカラムだけを抽出する
結合ノード	2つのテーブルを結合して1つのテーブルを作成する
集計ノード	指定したカラムの値の出現数を求める
合計ノード	指定したカラムの値別に、指定した数値カラムの合計を求める
相関ノード	指定したカラム間の相関係数を求める
関数ノード	数値カラムに対して関数演算を行う
ソートノード	指定したカラムをキーとしてレコードの順序を並びかえる
記述統計ノード	指定した数値カラム間の各種統計の処理を行う
スライスノード	指定したカラムについて同じ値を持つレコードをグループ化し、グループ毎にテーブルを分割する

「チャンネル別アクセス率」を求めるためには、「インタラクティブチャンネルへのアクセスデータテーブル」を使用する。このデータに、各サービス ID 別にサービス名を入れたカラム（カラム名：CH）を付加し、これをサービス名毎に集計し、結果テーブルを得る。DFD は図 3-1 のようになる。

ノードはクリックすることにより、演算に必要な条件を設定する画面をポップアップするようになっている（図 3-2）。例えば集計ノードでは、指定したカラムの値毎の出現数を数えることができる。具体的には「CH」を指定すると、チャンネル毎のアクセス数を数えることができる。出力されるテーブルは、サービス名とアクセス数の 2 カラムとなる。

またアークは DFD 実行後にアークアーク上のデータを表示することが可能となっている。そのため、DFD 検証も容易に行うことができる。



図3-1：DFD例

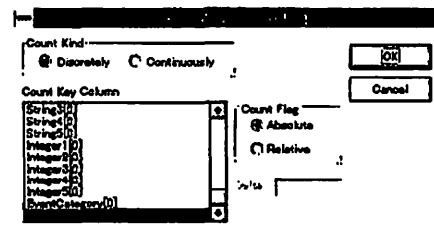


図3-2：集計ノードの設定例

図 3-1 の DFD を実行することにより「チャンネル別アクセス率」のデータテーブルを出力し、表計算ソフトを使用して図 3-3 に示すグラフを得る。

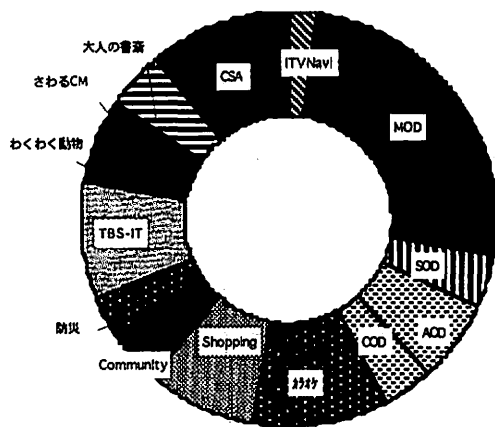


図3-3：チャンネル別アクセス率（'96.8）

図3-3のグラフは「チャンネル別アクセス率」であり、どのサービスがより利用されているかが把握できる。

これを見ると、「MOD:Movie On Demand」「カラオケ」の利用率が高い。「Shopping」の利用率もやや高いことも考え合わせると、映画、カラオケ、買い物という馴染みのものが、外出せずに行えるため、利用率が高くなる傾向にあると考えられる。

さらに、詳細に各サービスが利用されている

／利用されていない理由等を探るために、各サービス毎に、利用されているコンテンツやアクセスはしたが視聴するに至らなかったコンテンツを調べる等、推論を立てて実証することにもPAを利用できる。すなわち、分析方針に柔軟に対応できることは、PAの特徴といえる。

4. PA の適用例

・例1. 時間帯別アクセス／購買数

図4-1～4-3は「時間帯別アクセス数」を求めるためのDFDである。処理が複雑になるため、DFDを3つに分け、前のDFDの出力結果を次のDFDの入力に使用する。



図4-1:DFD1

図4-1のDFD1を実行する。ここでは「インタラクティブチャンネルへのアクセスデータテーブル」を使用する。新しいカラムを付加し、全ログについて1を入れる。これを集計することにより、ログの数を数えたことになる。

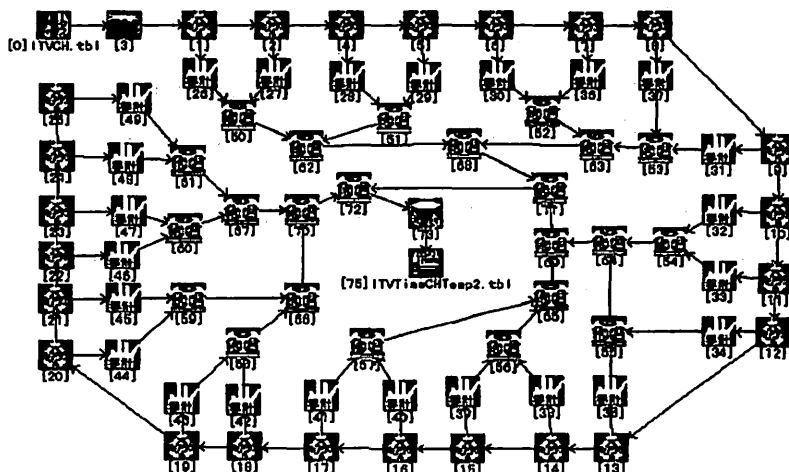


図4-2:DFD2

The diagram illustrates a video processing system architecture. It begins with an input block [0] TVTimeCHTemp.tb1, which feeds into a series of processing blocks. Block (2) is a circular block with a cross, followed by block (3) which is a square block with a cross. These feed into block (28), which is a square block with a cross, and then block (31), which is a square block with a cross. Block (31) feeds into a large grid of blocks. The grid consists of three rows of blocks. The first row contains blocks (4) through (13), the second row contains blocks (14) through (23), and the third row contains blocks (24) through (29). Each block in the grid is a square block with a cross. The output of the grid is block (30) TVTimeCH.tb1. The diagram shows a complex flow of data and processing steps, with various blocks representing different processing methods and channels.

最後に DFD3 を実行し、目的のテーブルを得る。ここでは、DFD1 と DFD2 の出力テーブルを結合し、全日数を入れたカラムを付加し、「(全アクセス数-各時間帯での null の数) / 全日数」の演算で、各時間帯の平均アクセス数を出す。この出力テーブルを表計算ソフトで処理したものが図 4.4 である。

ドが多いことに起因していると思われる。



全体にピークが低くなだらかなのは、データ解析期間が一般に夏休み期間であったため、平素よりも、平日でも在宅する人が多いためと思われる。



図 4-5 が示す時間帯別の月間平均購買数は、有料のサービスである映画、ドラマ、スポーツ、

アニメ、カラオケ等に対するものである。アクセス数に比例してはいない。つまり、アクセスするユーザが必ずしも購入するとは限らないわけである。アクセス数が1ピークを迎える10時台には、一旦コンテンツ一覧などを眺めてみるが購入までには至らず、午後になってから改めて購入する、という利用形態が想像できる。

・例2. 曜日毎アクセス/購買数

ここでは曜日別の場合の分析結果を示す。

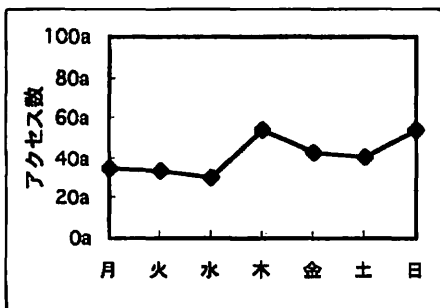


図4-6：曜日別アクセス数（'96.8）

図4-6は、曜日別のインタラクティブ・テレビへの月間平均アクセス数である。注目すべき点は、週末でも土曜日より日曜日のほうがアクセス数が多いことである。これは、土曜日は家族で外出し、日曜日は家でゆっくり過ごす人が多いとも考えられ、インタラクティブテレビの利用は、空き時間での利用が中心であり、通常のテレビと同様に受動的な使用状態であることがうかがえる。

ところで、平日であるにもかかわらず木曜日にアクセスのピークがあるのは特異的な現象だが、この日に新しいコンテンツが入ったことが原因と考えられる。

全体にアクセス数のばらつきが少ないのは、時間帯別の場合と同様に、夏休み期間で平素より平日の在宅人口が多いためと思われる。

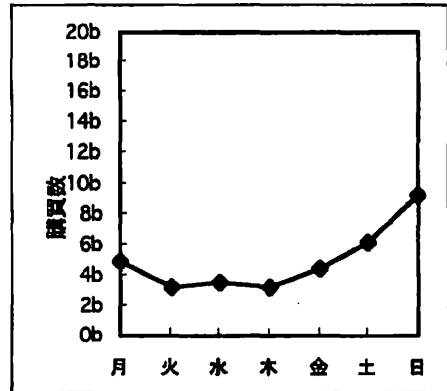


図4-7：曜日別購買数（'96.8）

図4-7は、曜日毎の月間の平均購買数である。木曜を除いてアクセス数にほぼ比例している。

5. おわりに

現在PAで実行できる分析は、サービス開始前に、結果を想定して作成している。このため実際にサービスを開始した今後、さらに詳しい分析や、新しい分析が必要となってくるだろう。PAはそのような要求に答えられるように設計しており、またその点が特徴である。今後は、既存の分析によるデータの解析や、新しい分析により、PAの有効性の実証を目標とする。

謝辞

本検討にあたり、NTTマルチメディアビジネス開発部の副島担当課長および関係者の方々に分析内容の検討の協力をいただきました。またNTTソフトウェア研究所第一プロジェクトの中村主幹研究員および関係者の方から貴重なコメントおよびご指導をいただきました。ここに深く感謝いたします。

参考文献

- [1] 大東敏信, 中西弘毅, 荒野高志. "Interactive SystemにおけるPreference Analysis Model". In 情報処理学会研究報告 95-DPS-70, pages 45-50, 1995.