

# Interactive System における Preference Analysis Model

犬束 敏信 中西 弘毅 荒野 高志

NTT ソフトウェア研究所

## 要旨

本論文では、インタラクティブ・システムに於いて、ユーザの振舞いを解析する、Preference Analysis Model を提案する。ユーザの振舞いを解析し、結果をフィードバックし利用することにより、インタラクティブ・サービスの質、効率を共に向上させることができる。ユーザの多様な振舞いに対して、効果的な解析を行なうためには、プログラマブルであることが望ましいが、それが訓練を要するほど複雑な作業ではオペレータの負担が大きく、また多大なコストが掛かるなどの問題が生じる。本モデルは、解析の対象・方法をデータフロー・ダイアグラムを用いて記述する。これにより、柔軟かつ容易に記述できることを示す。更に、VOD 管理システムに適用し、その有効性を確認している。

# Preference Analysis Model for Interactive System

Inuzuka Toshinobu Nakanishi Kouki Arano Takashi

NTT Software Laboratories

In this paper, authors propose Preference Analysis Model that analyzes behavior of the users in interactive system. Analyzing behavior of the users and feeding back the result to the system improve both quality and efficiency of service. To analyze efficiently miscellaneous behavior of the users, the description is desirable to be programmable. However if to program miscellaneous analysis is very hard and it requires the operator hard training, then the cost will be large. In this model, we describe analysis method by dataflow diagram, and it makes description easy and flexible. Authors ascertain the validity of this model applying to the VOD management system.

## 1 はじめに

On Demand Service のようなインタラクティブ・サービスでは、ユーザは受動的にサービスを受取るだけでなく、自らの意思に於いて能動的にシステム（サービス）への様々な働きかけを行なう。よって、その振舞いにはユーザの個性が反映されている。したがって、ユーザの振舞いを解析することにより個性を分析し、これを積極的に利用することによって、様々な良質・高効率なサービスを展開することが可能となる。以上から、インタラクティブ・システムにおけるシステムオペレーションでは、カスタマの情報管理、及びカスタマの動向をサービスへフィードバックする機能が重要な役割を占めることは明らかである。

本稿では、カスタマの振舞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 Pref-

erence Analyser と呼ぶ。Preference Analyser が持つべき機能の要件として、(A) 解析がプログラマブルに記述できること、なおかつ (B) 容易に記述・変更できること、が挙げられる。なぜなら、インタラクティブ・システムでは単一方向のサービスとは異なり、ユーザの振舞いがあらかじめ決まり切ったものではなく多種多様であり、すなわち、変化する（測定できる）パラメータも多種多様である。したがって、解析の方法・要求も多様であり、またそれらの変化のサイクルも短いと考えられる。このため、プロトタイプ的に短いサイクルで記述、解析が可能であること、要求の変化に迅速に対応できることが要求される。

従来は、GUI を用いてデータの統計処理が記述できる処理系としては、Excel や StatView のようなスプレッドシート形式のものが利用されてきた [3,4]。しかし、このような形式の記述では、

- 複数のテーブルのデータの関係から新しいテーブルが生成される場合、また、それが繰り返し行なわれる場合
- テーブルの要素（フィールド）に含まれるデータが更にテーブルになっている場合

などの複雑な処理は、(a) 記述ができない、あるいは (b) 複雑なマクロを定義する必要があり変更にかかるコストが大きい、さらに (c) マクロの記述には習熟が必要であり初心者には困難である、などの問題点があり、前述の要求を満たすことができない。

本稿で提案する Preference Analysis Model とは、前述の要件を満たすようにするために、データフロー・ダイアグラムを記述法として用いて Preference Analyser をモデル化したものである。本モデルはデータフロー・ダイアグラムを用いることによって、

- 柔軟性： オペレータにより、ランタイムにプログラミングが可能
- 簡易性： アイコンの組合せによる記述のため、データベース処理言語などの専門的知識が不必要
- 拡張性： ノード単位での追加・改版により機能拡張が可能
- 再利用性： 記述したものを部品として登録でき、これを用いた記述が可能
- 了解性： 階層的な記述により、視覚的な了解性が高い

という特徴を持っており、さらに、

- テーブルをアイコンとして扱い、その関係をノード・アークで記述するため、複数のテーブルを対象とする処理が容易に記述できる。
- テーブル内のフィールドの要素として、テーブルが含まれる場合でも扱える。

ことから、スプレッドシート形式による記述で問題点が解消される。

以下本稿では、まず解析対象であるデータの定義、アークの定義を述べる。さらに、ノードの詳細について記述を行ない、最後に、例を用いて本手法の有効性を示す。

## 2 データ属性

フィールドに格納されるデータの形式としては、数値、文字、テーブルの3種類の形式がある。また、数値、文字データはその値が連続値であるか、離散値であるか（量的データであるか、質的データであるか）により2つに分類が可能である。テーブル内のデータはこのうち、いずれかの形式・型を持つ。また、同一のカラムに含まれる全てのフィールドは同じ形式・型のデータが格納される。

本手法の特徴として、フィールド内にテーブルを置くことができる点があげられる。これにより、非単純定義域のデータを持つテーブルを正規化 [2] せずに扱うことが可能となる。

形式	型	備考
数値形式	整数	連続、離散 ※
	実数	連続
文字形式	文字列	離散
テーブル形式	ポインタ	

※整数形式は数学的には、離散値とみなすことができるが、解析時にはそのデータのとりうる要素の範囲が有限であるか無限であるかによって、離散データ、連続データの2種の場合に区別できる。言い換えれば、量を表すものが連続データ、名前・性質等を表すものが離散データである。

有限な集合：離散値データ

ex.) 月を示す整数値

↪ 文字形式に置き換え可（質的）

無限な集合：連続値データ

ex.) 料金を示す整数値

↪ 文字形式に置き換え不可（量的）

## 3 アーク

データフロー・グラフに於いて、ノード間の有向リンク、すなわちアークによりデータの依存関係を記述する。解析を開始するとアーク上を、データ（テーブル）が流れる。本モデルでは、データは世代を持たないものとする。したがって、サイクリックなアークの記述は許されない。

解析中、および解析終了後は、アークを指定することにより、そのアークを流れている、解析データ（カラム名のみ、またはデータ自身も表示、を選択可）を調べることができる。

#### 4 ノード

データフロー・グラフに於いて、ノードは(1)データ自身、(2)データに対する処理、(3)解析結果の出力方法、を定義する。すなわち、ノードには、データの流れの起点となるデータベース・ノード、データの処理を記述する演算ノード、および出力の形式を記述する出力ノードの3種類が存在する。また、記述したものを部品として登録し、これをマクロなノードとして使用することが可能である。

##### 4.1 データベース・ノード

解析の対象となる源データの集まりである。データはテーブル形式で提供され、一意に付けられたテーブル名でデータを表す。

また、解析後のテーブルを新しいデータベース・ノードとして登録し、再利用することもできる。

##### 4.2 演算ノード

演算ノードはデータまたは、データ群に対する操作を規定する。あらかじめ、提供される基本ノードの組合せにより複雑な処理を行なうマクロノードの作成・登録が可能である。現在、基本演算ノードには、以下に示す11種類を検討している。

1. Filtering (選択)
2. Categorizing (カテゴリ化)
3. Focusing (フォーカス)
4. Grouping (グループ化)
5. Joining (結合)
6. Expansion (展開)
7. Counting (集計)
8. Summarization (合計)
9. Correlation (相関)
10. Functional Operation (関数演算)
11. Sorting (ソート)

以下に、各々の演算ノードの機能を示す。

##### 4.2.1 選択

指定された検索条件に一致するレコードを取り出し、一致しないレコードは削除する。レコードの指定は単一でも、複数でもよい。

検索条件は、

カラムが文字形式の場合：

- ～を含む、～を含まない、～である、
- ～でない、～で始まる、～で終る、関数指定。

カラムが数値形式の場合：

- ～以上、～以下、～を越える、～未満、
- ～に一致、～でない、関数指定。

を利用できる。また、これら検索条件を複数指定した場合には、その関係をそれぞれ和集合、積集合とすることができる。

##### 4.2.2 カテゴリ化

指定されたカラムについて、同類に分類されるものをまとめ、タグを付与する。分類のための検索条件は選択ノードと同じものが選択できる。

以下に使用例を示す。

ID	Name	Tag1
0001	S. Young	QB
0002	J. Rice	WR
0003	J. Taylor	WR
0004	R. Watters	RB
0005	B. Jones	TE
0006	D. Sanders	CB
0007	K. Norton	LB

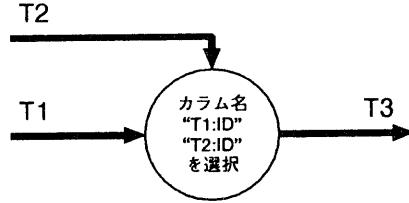


カラム"Tag1"に対し、  
OFF={WR, RB, TE, WR}  
DEF={LB, CB}  
を"Tag2"に付与する

ID	Name	Tag1	Tag2
0001	S. Young	QB	OFF
0002	J. Rice	WR	OFF
0003	J. Taylor	WR	OFF
0004	R. Watters	RB	OFF
0005	B. Jones	TE	OFF
0006	D. Sanders	CB	DEF
0007	K. Norton	LB	DEF

#### 4.2.3 フォーカス

指定されたカラム（カラム名を指定）を残して、残りのカラムを削除する。カラムの指定は複数を選択することが可能である。



#### 4.2.4 グループ化

同一サイズの複数のテーブルをグループ化／アングループ化する。テーブルをグループ化すると、ひとつのテーブルに対して操作を行なうと、残りのテーブルにも同様の操作が施される。

#### 4.2.5 結合

テーブル毎に選択されたカラムをキーとして、複数（3つ以上も可能）のテーブルを結合する。

結合されるテーブルは主従の関係を持ち、主となるテーブルの各レコードに対して、キーが一致するレコードが従となるテーブルに有るかどうかを調べ、一致するキーが有る場合には、該当するレコードへのポインタを、主となるテーブルのそのレコードに新たなカラムとして付加する。一致するキーが無い場合には、オペレータの指定により、nullを付加するか、またはレコードの削除を行なう。また、キーが一致するレコードが複数有る場合には、レコードのリストへのポインタを付加する。

T3

I D	Name	Tag1	Pointer1
0001	S. Young	QB	0xaaaa
0002	J. Rice	WR	0xbbbb
0003	J. Taylor	WR	null
0004	R. Watters	RB	0xcccc
0005	B. Jones	TE	null
0006	D. Sanders	CB	0xdddd
0007	K. Norton	LB	0xeeee

この結合において、nullとなるレコードを無視するようにオプションを選ぶと、ID = 3,5のレコードは削除される。

#### 4.2.6 展開

フィールド内のテーブルを展開し、テーブルを再構成する。結合時にキーとなったカラムは、重複しないように取り除く。

#### 4.2.7 集計

カテゴリ化されたテーブルに対して、または、離散データのカラムを含むテーブルに対して、タグあるいは、離散データについての要素別の出現頻度を調べる。

以下に使用例を示す。

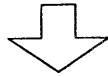
T1

I D	Name	Tag1
0001	S. Young	QB
0002	J. Rice	WR
0003	J. Taylor	WR
0004	R. Watters	RB
0005	B. Jones	TE
0006	D. Sanders	CB
0007	K. Norton	LB

T2

I D	Tel. Number	address
0001	03-4908-xxxx	0xaaaa
0002	03-4980-xxxx	0xbbbb
0004	03-4932-xxxx	0xcccc
0006	03-4921-xxxx	0xdddd
0007	03-4951-xxxx	0xeeee

I D	Name	Tag1	Tag2
0001	S. Young	QB	OFF
0002	J. Rice	WR	OFF
0003	J. Taylor	WR	OFF
0004	R. Watters	RB	OFF
0005	B. Jones	TE	OFF
0006	D. Sanders	CB	DEF
0007	K. Norton	LB	DEF



カラム"Tag2"  
を選択

Tag2	Frequency
OFF	5
DEF	2

#### 4.2.8 合計

カテゴリ化されたテーブルに対して、または、離散データのカラムを含むテーブルに対して、タグあるいは、離散データの要素別に、別カラムの数値データの和を求める

以下に使用例を示す。

ID	Name	Tag1	Tag2	TD
0001	S. Young	QB	OFF	5
0002	J. Rice	WR	OFF	34
0003	J. Taylor	WR	OFF	28
0004	R. Watters	RB	OFF	25
0005	B. Jones	TE	OFF	20
0006	D. Sanders	CB	DEF	8
0007	K. Norton	LB	DEF	3



カラム"Tag2"を  
キーに、TDを選択

Tag2	TD
OFF	112
DEF	11

#### 4.2.9 相関

複数のデータ群の相関（相関係数 [1,3]）を調べる。

また、相関だけでなく、一般的な統計処理も用意する。その代表としては、平均、分散等の記述統計が挙げられる [1,3]。

#### 4.2.10 関数演算

テーブル、カラム、フィールドに対して、各種の関数演算を行なう。

- 論理演算
- 算術演算
- 文字演算

#### 4.2.11 ソート

指定されたカラムの各フィールドをキーとして、各レコードをソートする。昇順、降順の選択が可能である。

#### 4.3 出力ノード

出力ノードは、その出力がどのように利用されるかを考慮して選択できるように、幾つかの出力方式を準備する。つまり、結果の出力はテキストだけでなく、以下のような5種類が利用可能である。

1. テキスト出力  
↳（標準的な出力）
2. グラフ出力  
↳（既存ツールを利用）
3. AP にコマンド送信  
↳（プログラムを生成）
4. テーブルをファイルに記録  
↳（DB アイコンとして追加）
5. マクロファイルに出力  
↳（メニューから選択し再利用）

また、アークをコピーして用いることにより、複数の方式に同時に出力することも可能である。

### 5 記述例

ここでは解析の一例として、『VOD管理システムにおいて、週末（金曜日）に良く見られた映画（3月分）のタイトル』を調べるための記述を紹介する。（図1）

VODサービスでの顧客の利用履歴を蓄積した Log DB (T1) と、コンテンツに関する情報を蓄積した Contents DB (T4) から、まず、「選択ノード」と「フォーカス・ノード」を用いて必要なデータだけを取りだし、それらを「結合」する (T7)。次にタイトル別に「カウント」し「ソート」すれば、上記の要求を満たす出力が得られる。この出力を VOD への広告としてフィードバックすることが考えられる。

- T1: 顧客ID、サービス内容、日付、プログラムID、…  
 ↓ 【日付、曜日、プログラムIDの列だけを取り出す】
- T2: 日付、曜日、プログラムID  
 ↓ 【日付が「3月」、曜日が「金曜日」のレコードを取り出す】
- T3: 日付、曜日、プログラムID
- T4: プログラムID、種別、タイトル、…  
 ↓ 【種別が「映画」のレコードを取り出す】
- T5: プログラムID、種別、タイトル、…  
 ↓ 【プログラムID、タイトルの列だけを取り出す】
- T6: プログラムID、タイトル
- ↓ 【T3とT6を結合する】
- T7: 日付、曜日、プログラムID、ポインタ(T6)  
 ↓ 【ポインタを展開する】
- T8: 日付、曜日、プログラムID、タイトル  
 ↓ 【タイトルの出現頻度を計算する】
- T9: タイトル、頻度  
 ↓ 【タイトルをキーにソートする】
- T10: タイトル、頻度

## 6 おわりに

本稿では、カスタマの振舞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持つ Preference Analyser の要件について整理し、それを踏まえて、容易にプログラムが可能で、複雑なデータ解析が可能なシステムとしてデータフロー・ダイアグラムを用いた Preference Analysis Model を提案した。また、本モデルの一部をプロトタイプとして作成し、予想される解析の要求を幾つかシミュレートし既にその有効性を確認している。

今後の課題として、記述したマクロ管理、記述の自動最適化等ツール類の充実と、それらを組み込んだアプリケーションの完成が挙げられる。

## 文献

- (1) 木下栄蔵:“多変量解析入門”, 啓学出版 (1987).
- (2) 原田勝, 今井恒雄, 平木茂子:“データベース構築の理論と実際”, コロナ社 (1985).
- (3) “Abacus Concepts, StatView Ver4.02”, Abacus Concepts, Inc. (1992).
- (4) “Microsoft Excel for Macintosh Ver4.0”, Microsoft Corp. (1992).

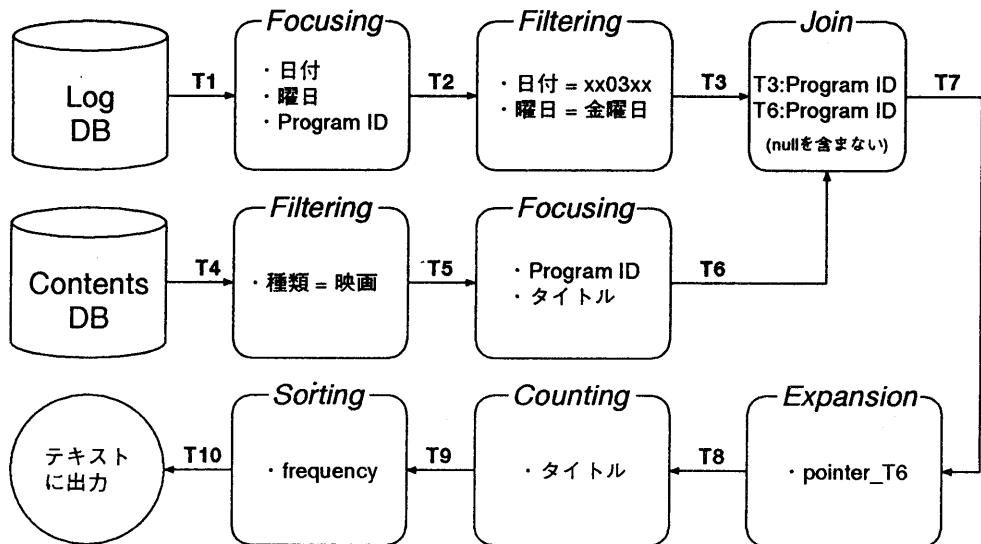


図 1: 記述例