社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2008－DBS－146 (56)
2008／9／23

履歴追跡型データモデルの評価

平井 規郎† 森山 令子† 郡 光則†

†三菱電機株式会社 情報技術総合研究所
〒247-8501 神奈川県鎌倉市大船 5-1-1
E-mail: †{Hirai.Norio , Moriyama.Ryoko , Kori.Mitsunori }@{dx,dr,ab}.MitsubishiElectric.co.jp

あらまし 近年多種多様な形式をもつ膨大なログがシステムから収集蓄積されるようになり，これらのログをただ蓄積するだけでなく，活用することが要求されている．従来ログの活用方法としては集計などのように全体の傾向に着目するマクロな視点にもとづいた方法が提案されてきた．しかしこの方法では追跡のように個々の人や物の動きに着目したミクロな視点での活用が困難であった．本報告では，ログに含まれる対象の前後関係を整理して管理することにより追跡を可能にする履歴追跡型データモデルを提案する．また提案手法を実装し，履歴追跡型データモデルの有効性および追跡効率を評価した結果について報告する．

キーワード データベース，データモデル，追跡

Evaluation of History Traceable Data Model

Norio HIRAI† Ryoko MORIYAMA† and Mitsunori KORI†

† Mitsubishi Electric Corporation Information Technology R&D
5-1-1 Ofuna, Kamakura, Kanagawa 247-8501, Japan
E-mail: †{Hirai.Norio , Moriyama.Ryoko , Kori.Mitsunori }@{dx,dr,ab}.MitsubishiElectric.co.jp

Abstract In late years , enormous log with various forms comes to be collected and accumulated by various systems, and what we utilize these log is demanded. Previously various methods have been suggested those command total tendency based on the macro viewpoint. However, these methods were inappropriate to utilize logs based on the micro viewpoint such as trace of history for each person or object. In this paper, we propose a History Traceable Data Model to enable a trace by managing the context of objects included in logs. And we report to evaluate effectiveness and efficiency of this data model.

Keyword Database，Data model，Traceability

1. はじめに

近年ストレージコストの低価格化および容量の増大を背景として，収集蓄積されるログの爆発的な増加や多様化に伴い，これらのログから有用な情報を抽出し，現象の分析や原因の特定に活用することが求められている．特に，セキュリティ，ビル／設備管理および製造トレーサビリティなどの分野ではログを用いて人や物を追跡することが求められている．しかしながら従来の RDBMS に代表されるデータベースによるログ管理では，項目によって構成された表を単位としてログを管理し，項目間および表間の関係を管理するものであるため，列間または表間の関係性にもとづくログ解析には優れているが，前後関係などの行間に存在する関係性を必要とする追跡などのログ解析は困難であった．そこでログを用いて対象を追跡するためにはログに含まれる関係性を管理するのに適したデータモデルを定義する必要がある．本報告では，上記データモデル（以下履歴追跡型データモデル）の定義について述べるとともに，本定義モデルを実装し評価した結果を示す．

2. 先行研究

収集・蓄積されるログの量が増大するにつれてログの処理時間も膨大になり，その結果ログは蓄積するのみで利活用がすすまないという状況があった．これに対して我々は，多様なログへの対応，高速蓄積・検索，ストレージ容量削減を目的として，大規模ログデータベースの開発をおこなってきた（文献[1],[2]）．その結果処理性能の向上により，従来困難であった大量のログに対する集計・検索処理が可能になった．一方ビル管理，製造およびセキュリティなどの分野では，従来の集計処理以外に，個々の人や物などを対象とする追跡などの分析ニーズが高まってきた．そこで我々は大規模ログデータベースを基盤として追跡を可能にするデータベースの研究に取り組んでいる．たとえば操作ログによる人の行動履歴の分析では，ユーザが操作したアプリケーションの時刻と位置情報を分類してユーザ操作をモデル化して分析するなどの研究があるが

（文献[3]），全体としての人の行動傾向を把握するものであって，個々の動きを対象とした分析ではない．トレーサビリティの観点では情報変化過程のトレーサビリティに関するデータモデルの研究（文献[4]）がある．この方法は情報の時系列変化を情報の変化差分に着目してモデル化したものであり，機器情報のトレースに有効である．しかし、一般的なログではなく数値データを対象としている点で課題がある．また分散する複数のデータソースにまたがって保持されるデータに対してトレーサビリティを実現するシステムに関する研究（文献[5]）では，統一的な検索手段を基盤とするトレーサビリティシステムの手法が提案されているが，データモデルの一般化に課題がある．

3. 履歴追跡型データモデル提案

本章では，まず本研究において使用する用語を定義し，次に我々が提案する履歴追跡型データモデル（文献[6],[7]）の定義について述べる．

3.1. 用語の定義

図 1 はあるユーザ（識別 ID=A）の入退室管理装置に記録されたログを関係データモデルおよび履歴追跡型データモデルで管理した場合のイメージ図である．図 1 をもとにそれぞれの用語について定義する．

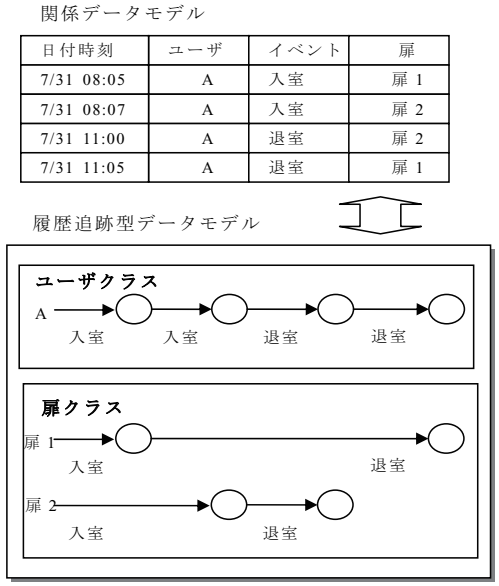

図 1　用語定義

3.1.1. クラス

追跡対象が属するカテゴリを表し，関係データモデルにおける項目に対応する．個々のクラスは全体において一意に識別可能なクラス ID をもつ．図 1 ではユーザおよび扉がクラスに該当する．

3.1.2. インスタンス

追跡対象の実体を表し，クラス内で一意に識別可能な ID をもつ．したがって，クラスの ID とインスタンス ID で追跡対象は全体において一意に識別できなければならない．図 1 では「A」「扉 1」「扉 2」がインスタンスに該当する．

3.1.3. イベント

入室，退室，ログインなど人や物の動作を表し，全体において一意に識別可能なイベント ID をもつ．インスタンスはイベントの前後で状態が遷移する．イベントは特殊なクラスであり，履歴追跡型データモデルでは必ず定義されていなければならない．図 1 では入室，退室がイベントに該当する．

3.2. 履歴追跡型データモデル定義

履歴追跡型データモデルとはイベントの発生前後における追跡対象の状態変化を対象ごとに整理して管理するためのデータモデルである。我々は追跡対象とイベントの関係構造を検討した結果，データモデルを定義する上で必要な関係構造は 3 つに分類できると考える．以下に履歴追跡型データモデルの定義に必要な 3 つの関係構造について説明する．

3.2.1. 順序関係構造

順序関係とは 1 つの追跡対象の状態をイベントによって関係付ける 1 対 1 の写像関係である．この写像関係はイベントの発生前後の状態を関係付ける写像関係であり，イベントの発生によって状態は遷移する．この状態とイベントの写像関係はグラフで表現することが可能である．図 2 において x_0，x_1，x_2 はインスタンス X の状態を表す．f_1 は状態 x_0 を x_1 に遷移させるイベント，f_2 は状態 x_1 を x_2 に遷移させるイベントである．図 2 の上図は写像関係で表現したもので下図はグラフで表現したものである．

3.2.2. 階層関係構造

階層関係構造とはあるクラスに属するインスタンスが同じクラスに属する異なるインスタンスを生成するようなイベントによって関係付けられる 1 対多または多対 1 の親子関係である．たとえばファイルをコピ

ーした場合は 1 つのファイルから複数のファイルが生成され，親子関係が発生する．またメールをリプライした場合はリプライによって元のメールとの間に親子関係が発生する．このように親子関係を発生させるイベントによって関係付けられる状態間の写像関係を階層関係と定義する．なお，異なるクラスに属するインスタンスを生成するイベントは考慮しないものとする．図 3 において x_0，x_1 はインスタンス X の状態，y_0 はインスタンス Y の状態である．X,Y はともに同一クラスに属するインスタンスである．またイベント f_1 は X が Y を生成する階層関係のイベントであり，このイベントにより x_0 から x_1 および y_1 に同時に遷移する．

3.2.3. クラス間関係構造

クラス間関係構造とは，異なるクラスに属するインスタンス間を関係付ける写像関係である．たとえばある時刻に記録された 1 つの入退室ログにはユーザが扉を通過したことが記録されている．この場合，履歴追跡型データモデルではユーザと扉の間の関係をクラス間関係構造として定義する．クラス間関係構造は RDBMS などで定義される項目間の関係と同等であり，1 つのログに複数のクラスが含まれる場合にそのクラスに属するインスタンス間はすべてクラス間関係構造によって関係付けられる． 図 4 において x_0,x_1,x_2 はクラス I に属するインスタンス X の状態を表し，y_0,y_1,y_2 はクラス II に属するインスタンス Y の状態を表す．イベント f_1 は x_0 と x_1 の状態を関係づける写像関係であるとともに x_0 と y_1 の状態を関係づける写像関係でもある．このとき図 4 の鎖線で表した関係がクラス間関係構造である．

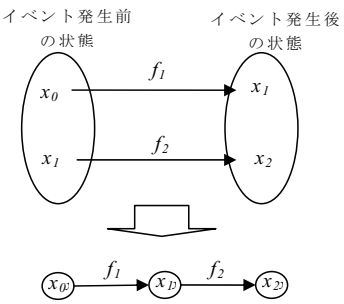

図 2 順序関係構造図

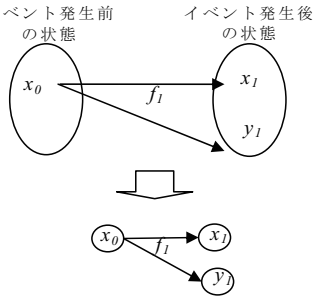

図 3 階層関係構造図

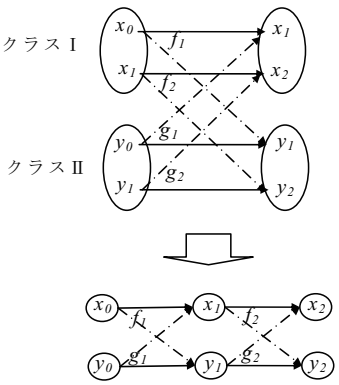

図 4 クラス間関係構造図

4. 実装

評価を目的として前章で定義した履歴追跡型データモデルを実装した．

4.1. 履歴追跡型ログ DB データ構造

履歴追跡型ログ DB は，ログから前章で定義した 3 つの関係構造を抽出し，対象ごとの関係を管理することにより効率的な追跡を可能にするためのインデックス（以下挙動インデックス）をメモリ上に生成する．挙動インデックスはログから 3 つの関係構造を抽出しグラフ構造として管理する．履歴追跡型データモデルを実現する挙動インデックスのデータ構造図を図 5 に示す．挙動インデックスが管理する主なデータ構造と

しては,クラス情報構造体,インスタンス情報構造体,階層情報構造体，順序履歴情報構造体がある．クラス情報はインスタンス情報を管理する．階層情報および順序履歴情報は個々のインスタンスによって管理される．階層情報は階層関係を発生させたイベント，子のインスタンスおよび対応する順序履歴へのポインタを管理する（図 5 点線）．順序履歴は属するインスタンスで発生したイベントの発生順に格納され，前後関係のポインタとともに管理される．クラス間関係構造は異なるクラス間に属する順序履歴同士を関係づけるポインタとして管理される（図 5 の鎖線）．

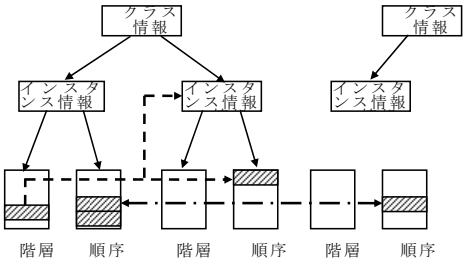

図 5　実装データ構造

4.2. 履歴追跡型ログ DB システム構成

履歴追跡型ログ DB は，履歴追跡型データモデルを実現する挙動インデックスを生成管理する管理部，追跡処理を実行する実行部および追跡結果をグラフ表現で表示する結果表示部から構成される．履歴追跡型ログ DB として実装したシステム構成図を図 6 に示す．図 6 の標準形式データとは挙動インデックス生成のためにログを共通の形式に変換したデータである．

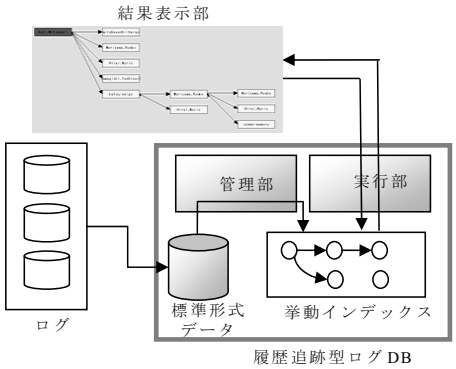

図 6　履歴追跡型ログ DB システム構成図

5. 評価実験

本研究の目的は定義した履歴追跡型データモデルがログ情報の追跡に有効であること，および一定の追跡性能を満たすことを確認することにある．プロトタイプであるため，必ずしもメモリ性能および処理性能は最適化されていないが，評価した結果について以下に述べる．

5.1. モデルの有効性評価

本研究で定義した履歴追跡型データモデルの有効性評価の目的は 3 章で定義した順序関係構造，階層関係構造，クラス間関係構造から構成される履歴追跡型データモデルにもとづきログを管理することにより，すべての情報の追跡が可能であるかどうかを検証することにある．そこで，本研究では生成した実験データおよび実データを適用しモデルの有効性評価を行った．本稿では適用したデータから以下の 2 つのデータに対して評価した結果について述べる．

① シミュレーションデータ
② メール送受信履歴

5.1.1. シミュレーションデータによる評価

定義した 3 つの関係構造をもつ単純な評価用シミュレーションデータを生成し適用した．以下に生成したデータのデータモデル図を示す．なおクラスは 2 つあり,個々のクラスには 2 つのインスタンスが存在する．クラス 1 では階層関係イベントによってインスタンス 1 がインスタンス 2 を生成する．

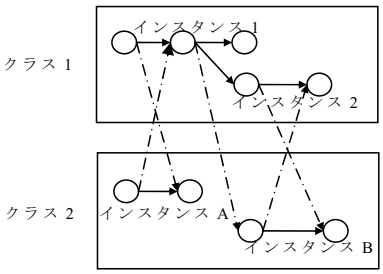

図 7　評価用シミュレーションデータ

追跡結果 1

クラス 1 に属するインスタンス 1 を追跡した結果を図 8 に示す．図 8 よりクラス 1 の順序関係構造および階層関係構造を管理することにより追跡可能であることがわかる．

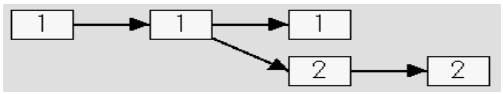

図 8　階層関係追跡結果

追跡結果 2

クラス間関係構造を用いてクラス 1 に属するインスタンス 1 の動きと関係するクラス 2 のインスタンスを追跡した結果を図 9 に示す．図 9 よりクラス 1 のインスタンス 1 の関係をたどることによりクラス 2 の「インスタンス A」と「インスタンス B」の追跡が可能であることがわかる．

図 9　クラス間関係追跡結果

5.1.2. メール送受信ログによる評価実験結果

メール送受信ログとはメールのヘッダ情報である．メール送受信ログ追跡ではヘッダ情報に含まれる時刻，送受信アドレス（From, To, Cc など），メッセージ ID（メールを一意に識別する ID），In-Reply-To（返信の親メールのメッセージ ID）等の情報から関係構造をもつ挙動インデックスを生成管理する．メール送受信ログのデータモデルはユーザクラスとメッセージ ID クラスの 2 つのクラスをもつ．メール送受信ログ追跡では以下の追跡を行い評価した．

- メール配信経路の追跡
- ユーザ送受信履歴の追跡

メール送受信ログにおけるデータモデル図の例を以下に簡単に示す．図 10 は，たとえば 2 人のユーザ A およびユーザ B の間でのみ発生したメールのやりとりを履歴追跡型データモデルで表現したものである．

メールの配信経路の追跡を関係データモデルにより追跡する場合には，以下の①～⑤の処理を返信回数分繰り返す．

① 追跡対象メッセージ ID を指定
② 指定したメッセージの受信者を取得
③ 指定したメッセージ ID を親に持つ子メッセージ ID を取得
④ 子メッセージ ID を受信したユーザを取得
⑤ 親メッセージ ID 受信者と子メッセージ ID 受信者の関係を生成

これに対して履歴追跡型データモデルでは，ユーザの履歴とメッセージ ID の履歴の他にクラス間関係構造でメッセージ ID とユーザが図 10 のように関係付けられているため，追跡対象となるメッセージ ID を 1 度指定すれば関係構造をたどることですべての配信経路を取得することが可能である．

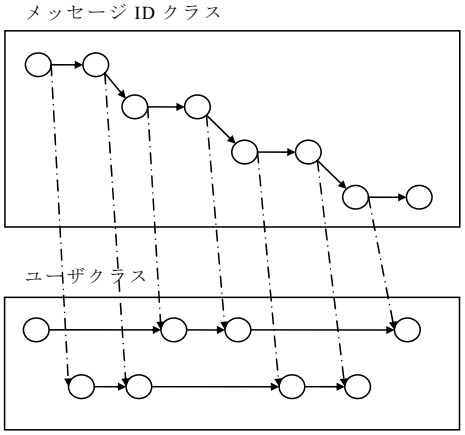

図 10　メール送受信ログデータモデル図

実装したシステムに実際のメール送受信ログを適用して追跡した結果，すべてのメールおよびユーザについて追跡が可能であることを確認した．図 11 にあるメールの配信経路を追跡した結果例を示す．

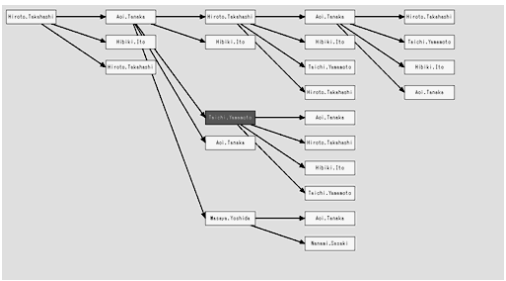

図 11　実メール送受信ログ追跡結果例

5.2. 性能評価実験

本節では性能評価の結果について述べる．本報告における性能評価の目的は，我々が定義した履歴追跡型データモデルに実際のログを適用し，対象の追跡を実時間内に処理することが可能かどうかを確認することである．そこで実装したシステムに実際のメールヘッダ情報から生成したメール送受信ログを適用し，追跡性能を評価した．

本評価で用いた性能評価環境を表 1 に示す.

表 1　性能評価環境

OS	Microsoft(R) WindowsXP Service Pack3
CPU	Intel(R) Core(TM)2
メモリ	1.98GB RAM
HDD	149GB

性能評価結果

実際のメール送受信ログを適用し，評価した結果について示す.

モデルの有効性評価でも述べたように関係データモデルによるメール配信経路の追跡では，返信が発生したかどうかを確認するために，すべてのメッセージについて親子関係があるかどうかを調べる必要があり，計算量が膨大である．さらに取得したメッセージ ID をもとに関係する送受信者を取得し，関係を生成するなどの処理も含めると追跡処理には不向きである．一方履歴追跡型データモデルではメッセージ ID 間の親子関係，およびメッセージ ID とユーザの関係をデータモデルとして管理しているため，効率よく追跡することが可能である．そこで入手可能なメールログを対象として測定した結果，メッセージ数またはユーザ数に依存することなく同等の測定結果が得られた．メール配信経路の取得について測定した結果を表 3 に示す.

表 3　メール配信経路取得時間

返信数	計測時間（ミリ秒）
2	2
3	2
4	4
5	5
8	5
10	5

6. 考察

モデルの有効性評価では，メール送受信ログをはじめとしていくつかの実データを適用し追跡した結果，我々の定義した 3 つの関係構造をもつ履歴追跡型データモデルによりすべての対象に対して追跡可能であることを確認した．また性能評価では表 3 からわかるように，従来のデータモデルでは非効率的な処理であったメール配信経路の取得を効率よく実時間内で処理することが可能であることを確認した.

7. まとめと今後の課題

本研究ではまず対象の追跡を可能にするためにログのもつ情報からイベントと対象の関係構造を定義し，これらの関係構造によって構成されるデータモデルを履歴追跡型データモデルとして定義した．次に定義した履歴追跡型データモデルを実装し，モデルの有効性および性能について評価を実施した．その結果以下の結論を得た.

- 順序履歴および階層履歴についてイベントの発生順または逆順に追跡することが可能
- 追跡対象と関係する異なるクラスの追跡が可能
- 従来は困難であった追跡処理を実時間内で処理することが可能

今後は性能向上とともに，履歴追跡型ログ DB における検索についてクエリ言語を含めた検索方式の検討をすすめていく.

文　献

[1] 竹内　丈志，山岸　義徳，中村　隆顕，郡　光則：“大規模ログデータベースの評価”，IPSJ68 回全国大会講演論文集，ID-1，2006.

[2]中村　隆顕，山岸　義徳，竹内　丈志，郡　光則：“大規模ログデータベースの実現”，IPSJ68 回全国大会講演論文集，ID-2，2006.

[3]松本　光弘，清原　良三，福井　秀徳，沼尾　正行，栗原　聡:“携帯端末による人物行動履歴の分析に関する一考察”，第 21 回人工知能学会全国大会講演論文集，2F3-3，2007.

[4]大木　康幸，有澤　博：“ユビキタス環境における情報変化過程のトレーサビリティのためのデータベースモデル”，DEWS2005.

[5]百合山　まどか，小金山　美賀，渡邊　裕治，北山　文彦，沼尾　雅之：“疎結合な関係にある企業間のトレーサビリティシステムの提案”，DEWS2006.

[6]平井　規郎，森山　令子，郡　光則：“履歴追跡に適応するデータモデルの検討”，IPSJ70 回全国大会講演論文集，3B-6，2007.

[7]森山　令子，郡　光則，平井　規郎：“履歴追跡結果の表示方式の検討”，IPSJ70 回全国大会講演論文集,3B-7，2007.