履歴追跡に適応するデータモデルの検討[1]

平井規郎[2]　森山令子[3]　郡光則[4]
三菱電機株式会社　情報技術総合研究所[5]

1. はじめに

近年ストレージ価格の低下を背景に、様々な情報システムから生成される多様なログの蓄積・保存がすすんでいる。その結果ログに対する利活用の要望が高まっている。しかしログの形式はデータソースごとに異なるため、同じ追跡対象が複数のログに記録されているような場合、追跡対象の相互関係を把握しながらイベントの発生した順番に追跡することは困難である。

本稿では、多様な形式を持つ複数のログに記録された追跡対象ごとにイベントの発生による状態遷移をグラフ構造で表現し効率よく追跡可能なデータモデルの提案を行う。

2. 関係データモデルの課題

表1はPC操作ログを関係データモデルで表現した一例である。表1のログではユーザAが「ログイン」「Iを開く」「IをIIにコピー」の操作履歴である。

表1 PC操作ログ例

タイムスタンプ	操作	ユーザ	ファイル1	ファイル2
hh:mm:ss	ログイン	A	−	−
hh:mm:ss	開く	A	I	−
hh:mm:ss	コピー	A	I	II

一般的に関係データモデルに格納されたログは表1のように、イベントが発生した時刻やイベント内容（表1の例では操作）以外に複数の項目(ユーザ、ファイル1、ファイル2など)に関する情報を含む。ここで追跡対象とはユーザ、ファイル1、ファイル2などの各項目の要素（たとえばA、I、II）である。関係データモデルでは表同士の関係および行方向での関係については効率よく管理できるが、履歴の発生順序を管理するための列方向関係の管理は困難であり、その点が履歴追跡システムへの関係データモデル適用の課題である。

1 DataModel for Traceable Log Database. 2 Norio Hirai 3 Ryoko Moriyama 4 Mitsunori Kori 5 MITSUBISHI ELECTRIC CORPORATION INFORMATION TECHNOLOGY R&D CENTER

3. 履歴追跡型データモデルの提案

本稿では関係データモデルでは困難な履歴追跡に適応するデータモデルを提案する。

3.1. グラフ構造による表現

追跡対象のふるまいを発生順に関係づけて表現するためには、グラフというデータ構造を用い表現が有効である。本節ではグラフ構造を用いて、追跡対象とイベントの関係を統合して表現する。

3.2. イベントと追跡対象の関係

従来の追跡用システムでは追跡対象に対して一意のIDを付加することで追跡を可能にする方法が提案されてきた。この方法ではファイルの追跡を例にすると表1の「I」を追跡する場合、コピー操作が行われた時点で「I」と「II」が同じファイルであることを示すIDをシステムが付加する。しかし、ID管理により追跡対象の一意性を確保する方法では、メールの送受信、特定できない人の動き、原料の統合・分割などにおける追跡対象の複雑な動きを表現することが困難である。そこで、本提案モデルではログに記録されるイベントの発生前後における追跡対象の関係をグラフ構造で管理する。

今、追跡対象のイベント発生前の状態をA、イベント発生後の状態をB、またそのとき発生したイベントをfとすると$B=f(A)$と表すことができる（図1）。これはA、Bをノード、イベントを有向辺とするグラフである。

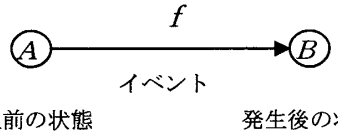

図1　イベントと追跡対象の関係概念図

ある追跡対象の状態集合$X=\{x_0,x_1,\cdots,x_{n-1}\}$（添え字は発生順で追跡対象に分岐はないものとする）に対してイベント集合$f=\{f_1,f_2,\cdots,f_n\}$が発生したとすると、同様に以下の関係が成り立つ。

$$x_1 = f_1(x_0), \cdots, x_n = f_n(x_{n-1})$$

したがって、追跡対象の初期状態 x_0 から x_n までをイベント発生順に追跡することはつまり、イベント集合 f および合成演算子「∘」　を用いることにより

$$x_n = f_n \circ f_{n-1} \circ \cdots f_2 \circ f_1(x_0)$$

と表すことができる。また追跡対象が分岐した場合においても分岐前と分岐後の前後関係を管理することで同様に表すことができる。つまり、イベントによる追跡対象の前後関係と発生順序を管理しグラフ構造で表現することにより、すべての状態をイベント発生順（または逆順）に追跡することが可能になる。

3.3. 履歴追跡型データモデル

前節で追跡対象とイベントの関係をグラフデータ構造を用いて表現することで、追跡が可能になることを説明した。本節では追跡対象とイベントの関係を２種類のパターンに分類し、分類ごとに異なるデータ構造を管理することにより効率よく追跡可能な履歴追跡型データモデルを提案する。

履歴追跡型データモデルは追跡対象を主体とするデータモデルである。追跡対象はクラスとインスタンスにより一意に特定される。クラスとは関係モデルにおける項目にあたるものであり、たとえば「ファイル」「ユーザ」などである。またインスタンスは各クラスに含まれる要素であり表 1 の「A」「I」などがこれに該当する。

○クラス内の関係

同じクラス内での基本的な関係構造としてはイベントの前後でインスタンスの状態が 1 対 1 に対応する順序関係（図 2）、1 対多または多対 1 の階層関係（図 3）の２つの関係構造を定義する。たとえばファイルがコピーで複製される場合は図 3 で表される階層関係構造をもつ。

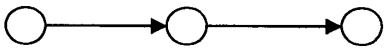

図 2 順序関係構造図

図 3 階層関係構造図

○クラス間の関係

異なるクラスに属するインスタンスの状態間に発生する関係がクラス間の関係である。たとえばあるファイルに対してあるユーザが編集操作を行った場合、その編集操作に対してユーザクラスのインスタンスとファイルクラスのインスタンスがクラス間関係を持つ。この関係をクラス間関係構造として定義する。

○モデル適用例

上記２つの関係構造をもつデータモデルを履歴追型データモデルとよぶ。表 1 の履歴データをもとに履歴追跡型データモデルの生成例を示す（図 4）。図 4 において実線が順序関係、破線が階層関係、1 点鎖線がクラス間関係である。ユーザＡは「ログイン」「開く」「コピー」で順序関係を持つ。またファイルクラスはＩが「開く」「コピー」で順序関係を持ち、ⅡがＩと「コピー」で階層関係を持つ。また、ＡとＩは「開く」および「コピー」でクラス間関係をもつ。

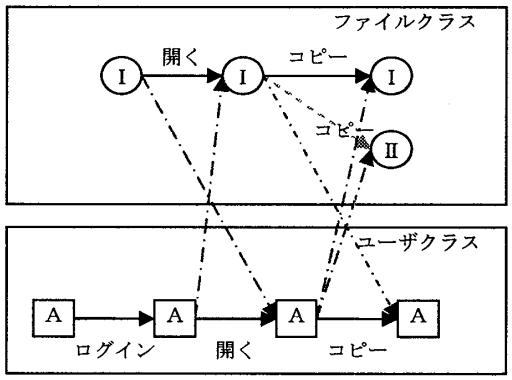

図 4 履歴追跡型データモデル概念図

上記履歴追跡型データモデルにおいて各クラス内でのインスタンスの追跡が可能になる。また追跡の途中でクラスを切り替えて（たとえばユーザからファイル、ファイルからユーザ）追跡することも可能になる。

4．おわりに

本稿では、履歴データから効率的に追跡を行うために必要な履歴追跡型データモデルを提案した。今後は本提案モデルを実装し評価を行なう予定である。

参考文献

[1] 森山，平井，郡，“履歴追跡結果の表示方式の検討”，IPSJ 70 回全国大会予稿集