

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 28 日現在

機関番号：12612

研究種目：基盤研究（C）

研究期間：2009 ～ 2011

課題番号：21500132

研究課題名（和文） トレーサビリティ・データマイニングの研究

研究課題名（英文） A Study on Traceability Data-Mining

研究代表者

沼尾 雅之（NUMAO MASAYUKI）

電気通信大学大学院・情報理工学研究科・教授

研究者番号：90508821

研究成果の概要（和文）：

トレーサビリティ・データマイニングとは、自動車産業のように数万点の部品からなる製品を、複数の部品メーカーが階層的に組み立てる過程において、製品の故障が、どの部品に起因するかを解析するための技術である。本研究では、仮想ロット分割法を導入し、故障の可能性の高い部品だけを選択的に探索できるリルダウンマイニング法を開発した。実装、評価した結果、従来法に比べ、データサイズが増加しても探索時間が短く、実用的であることを示した。

研究成果の概要（英文）：

Traceability data-mining is a technology to analyze which of the parts contributes to the product-level fault in the assembly processes such as automobile manufacturing, where more than ten-thousands of parts are assembled by multiple vendors in hierarchical manner. In this study, we introduced a virtual lot concept and developed the drill-down mining method which can search the suspicious parts selectively. After implementing the system, we compared our method with the existing one and showed that our method has a feasible search time even if data size is increased.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,700,000	510,000	2,210,000
2010年度	1,100,000	330,000	1,430,000
2011年度	600,000	180,000	780,000
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：知識発見とデータマイニング

1. 研究開始当初の背景

(1) 階層型データマイニング技術

トランザクション系 DB からのデータマイニングについては、売り上げ（POS）デー

タからバスケット分析によって、購買品の相関関係を求めるような多くの研究がある。その中でも重要なものが、分散並列データマイニングであり、2つのタイプに分けられる。

①単一データソースからの相関関係を、マルチプロセッサを用いて計算させるものと、②分散データソースから全体の相関関係を解析するものである。②はさらに、同種のトランザクションが分割されたデータを対象にする水平分割のものと、アイテムの種類ごとに分割されたデータを対象にする垂直分割のものに分類される。

本研究で対象とするデータは分散階層型データであり、垂直かつ水平のデータ分割が、階層的に行われているという点で、従来の分散データマイニング手法では扱えないものである。しかし、この分散階層型データは、製造物のトレーサビリティ・データとしては非常に一般的なものである。たとえば、自働車では、エンジン、シャーシなどの最終部品から構成されており、これがトップレベルの分割であるが、それぞれの最終部品は、たとえばエンジンはシリンダーとコンロッドから構成されているなど、第2レベルの分割があり、これが繰り返される。さらに、自働車の最終組み立ては自動車メーカーで行われるので、組み付けデータはメーカーに保管されるが、エンジンの組み立ては別のメーカーで行われ、そのデータは別のメーカーに保管される。このように、データがPart-Of階層関係にほぼ等しい形で分割されているのが、トレーサビリティ・データにおける分散階層型データの特徴である。

我々は、こうした特徴を持つ分散階層型データに対する収集・検索方法や、多次元データ解析やデータマイニング手法について研究している。多次元データ解析(OLAP)では、故障の原因追跡の手段として、故障に最も貢献している属性を、その属性と故障の共起関係によって調べる。ところが属性は部品ごとに異なるため、最終製品で起こる故障と、部品の属性との共起関係を調べるためには、部品テーブルにドリルダウンした状態で故障頻度との相関を見なければならぬ。われわれは、この部品レベルの属性と製品レベルの故障の間の相関を、効果的に取得する方法として、ロットに着目することを提案している。これは、ロットは部品製造の単位であり、1つのロットの中の部品は、ほぼ同じ属性値を持つことが仮定できるからである。

(2)トレーサビリティ技術

トレーサビリティとは、RFIDなどから得られるセンサーデータを基にして、人・物・情報の流れを可視化し、解析するためのものである。特に製造工程においては、原材料メーカーから最終製品メーカーまで複数の業者が関与し、それぞれの段階で部品の製造・組み立てが行われる。したがって、ある部品は、別の部品に組み付けながら、複数の業者の間を移動していくことになり、この履歴を検索するのがトレーサビリティである。トレーサ

ビリティには、製品からその部品を製造した業者に遡っていくバックワード・トレーサビリティと、部品からその部品が組みつけた製品を検索するフォワード・トレーサビリティがある。我々はこれまで、リコールプロセスには、この2つのトレーサビリティの連携が必要であり、そのための企業間データ統合について研究して、論文発表をしている。

製造工程における各段階の業者は、下工程の業者から来た部品を加工・組み付けて、上工程の業者に供給していく。この時に、下工程の部品IDと上工程の部品IDのマッピングが行われるが、このマッピングテーブルを有効に管理しておくことが、後のトレーサビリティに必要である。我々は、このテーブルを最終製品ごとに集計し、これをインデックスとして管理する方法を提案した。これを生成するには、製品が作られるたびに部品すべてのテーブルを走査することが必要であるが、この情報を製品に付加しておくことによって、最終メーカー側で簡単にインデックスの更新ができるようにした。

2. 研究の目的

本研究の目的は、大量かつ分散して発生するトレーサビリティ・データから、知識発見をするためのデータ・マイニング技術を開発するものである。トレーサビリティ・データとは、たとえば製造工程における部品の組み付け時に、RFIDによって検出されるデータのように、各工程で機械的に大量に生成される特徴を持つ。また、原材料工場、部品製造工場から製品組み立て工場など、製造プロセスに応じて部品が移動するため、1つの製品ID・部品IDを持つデータの保管場所も複数になる。

本研究では、この分散トレーサビリティ・データが階層構成を持ち、それが、製品の階層構成(BOM: Bill of Material)と似ていることに着目したドリルダウン型データマイニング手法を提案し、実用化に向けた技術開発をすることを目的とする。これは、最近問題になっている、製造業における品質問題を解決するためのキーテクノロジーとなるものである。たとえば、自動車産業では、1台当たり数万点の部品から構成され、それぞれの部品が異なる業者から供給されている。このため、故障原因の究明が難しくなっており、それがリコールの遅れなどにつながっている。本研究の手法により、故障が出始めた段階で、その可能性となる部品を特定することが可能になるので、早期の問題発見(Early Warning)が可能となる。

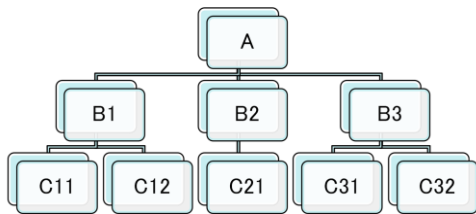
3. 研究の方法

本研究は、トレーサビリティ・データマイニングのアルゴリズム設計と計算機実験の

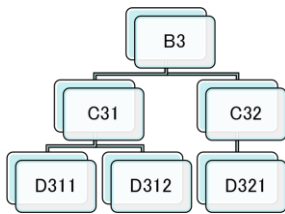
2つから構成される。研究期間の前半には、アルゴリズム設計と評価用プログラムの開発を中心に進め、後半には、分散階層データ環境における計算機実験と、トレーサビリティ・データを実際に生成した実証実験を行い、実用性を含めた総合的な評価をする。

(1) 製造物トレーサビリティ・データのモデル化

対象となるトレーサビリティ・データの特徴を明らかにする。部品の階層構造は、図1のようなPart-Of 関係になる。製品・部品の組み付けデータは、1つの大きな木として管理されるのではなく、例えば図1(1)(2)のように、それぞれのノードがその1つ下の階層分だけを保持した分散データ構造とな



(1) 製品 Part-Of 階層



(2) 部品 Part-Of 階層

図1. ドリルダウン型マイニング

る。このようなデータ構造をXMLなどでモデル化することによって、サーバ間で交換可能な標準データ形式を設計する。

(2) 分散階層構成を持つデータに対するマイニングアルゴリズムの開発

分散階層データ構造に適合したドリルダウン型のマイニングアルゴリズムを設計する。図1(1)のような製品レベルでの属性値(たとえば故障)と、1つ下のレベルの部品属性との相関関係から、製品レベルでの属性値に最も寄与している部品を検索する。例えば製品Aを構成する部品B3が故障に関与している可能性が高いとする。そこで、今度は、図1(2)のように、部品B3を構成しているサブ部品との関係を調べていく。これを繰り返すことによってドリルダウン型マイニングが可能になる。

(3) ドリルダウン型マイニングプログラムの実装と評価

ドリルダウン型データマイニングでは、すべてのデータを一か所に集めたバスケット

分析に比べ、もし、疑わしい部品の階層だけを発見できるとすると、対象となる属性の数は、 $O(N \cdot M)$ から $O(N \cdot \log M)$ となり、大幅な効率の向上が図れる。ここで、 N は部品数、 M は部品あたりの属性数の平均値である。そのためには、事前に計算されたロット情報との相関によって、疑わしい部品が選択できなければならない。したがって、従来型アルゴリズムとは、事前計算と相関計算の効率の2つの面から比較・検討していく。

(4) 製造工程トレーサビリティ・データによる故障原因分析の評価

実際の製造工程を模倣した実験環境を構築し、RFIDタグとリーダーを利用することによって、製造工程における組み付けデータを収集し、各製造者のデータサーバに保管し、これらのデータサーバをネットワークで接続することによって、本研究で開発したドリルダウン型マイニングシステムを実証実験する。

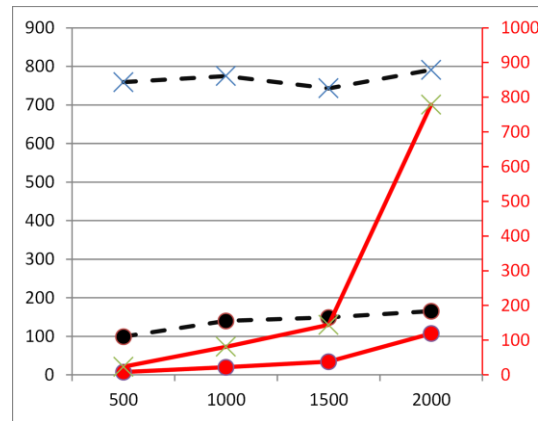


図2. アルゴリズムの評価

4. 研究成果

以下のような研究成果が得られた。

(1) 製造物トレーサビリティ・データのモデル化

製品の階層構成(BOM: Bill of Material)は、製品設計と製造の両方に用いられる。このBOM構造を、データマイニングが対象としている分散階層構成を持つデータモデルの1つとして定義した。さらに、部品製造の単位であるロットを、マイニングの属性値として使えるようにモデルを拡張した。

(2) 分散階層構成を持つデータに対するマイニングアルゴリズムの開発

複数製品にわたる故障データから、効率的にその原因となった部品、その製造者、およびそのロット番号を特定するためのドリルダウン型マイニングアルゴリズムを開発した。ドリルダウン型マイニングを可能にするには、各階層で、組み付けられる部品の1つ下の部品情報が必要となる。例えば、図1(1)

では A を構成する B1, B2, B3 の情報は得られるが、さらに C11, C12, C21, C31, C32 の組み付け情報が必要だが、これを簡単に実現する方法として、ロット細分割法を考案した。そして、組み付け時にロット分割情報を下位から上位の階層に持ち上げ、マイニング時には上位から下位へドリルダウンするため実行制御法を開発した。

(3) ドリルダウン型マイニングアルゴリズムの実装と評価

ドリルダウン型マイニングの実行制御アルゴリズムを、実行可能なプログラムとなるように、Java を用いて実装した。そして、分散計算機環境下でこのプログラムと既存の並列分散マイニングアルゴリズムを実行させて比較評価を行った。さらに、ネットワークを介した分散アルゴリズムの評価も行った。評価は、計算時間、探索対象属性数、データソースアクセス数などを、製造工程のさまざまな前提条件下で仮想的にデータを生成して行った。このために、仮想的なトレーサビリティ・データを自動的に生成プログラムも開発・実装した。図 2 は、データサイズに対して提案法 (○プロット) と従来法 (×プロット) の DB アクセス数 (破線・左軸) と探索時間 (実線・右軸・単位秒) を比較したものである。提案法はデータサイズの増加に対して特に有効であることが示された。

(4) 製造工程のトレーサビリティ・データにおける故障原因分析の評価

実際の製造ラインを想定したデータ生成と故障原因分析への応用可能性について実験した。実験環境は、RFID によって部品受領と製品出荷のタイミングを取得できるようにし、データは各ノードのサーバで保管し、インターネットの標準 WebServices プロトコルである EPCIS によってネットワークを介してデータ交換ができるようにした。さまざまな製造条件と不具合条件の下で実データを収集して、提案マイニング手法を既存方法と比較できるような環境を構築した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 3 件)

- ① 松澤裕史, 沼尾雅之: 集約バスケットからの相関関係マイニング, 電子情報通信学会論文誌 D (査読有り), Vol. J95-D, No. 2, pp. 170-182, 2012
- ② 沼尾雅之, 松尾総一郎: プロセス産業のための履歴テーブルに基づく品質分析法の提案, 日本データベース学会論文誌 (査読有り) Vol. 10, No. 1, pp. 79-84, 2012.
- ③ Y. Kuniyoshi and M. Numao: Academic Roadmap in Integrated Information

Field, Technical Note, Advanced Robotics (査読有り), Vol. 23, No. 11, pp. 1465-1474, 2009

[学会発表] (計 10 件)

- ① 中野 隆介, 沼尾 雅之: 無線 LAN アクセスポイントへの検索要求を利用した鉄道車内混雑度推定, 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012), 2012 年 3 月 5 日, 舞子ビラ神戸 (兵庫)
- ② 高橋 麻美, 根路銘 崇, 松澤 裕史, 沼尾 雅之: 利用者単位での消費電力可視化システムの開発, 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012), 2012 年 3 月 3 日, 舞子ビラ神戸 (兵庫)
- ③ 根路銘崇, 高橋麻美, 松澤博史, 沼尾 雅之: サービス消費モデルを用いて要求を満たせる電力消費最適化手法の提案, 情報処理学会第 174 回ソフトウェア工学研究会, 2011 年 11 月 2 日, 奈良県新公会堂 (奈良)
- ④ 高橋麻美, 根路銘崇, 沼尾 雅之: サービスインターフェースモデルに基づく利用者単位の消費電力の測定手法, 第 10 回情報科学技術フォーラム (査読有り), 2011 年 9 月 7 日, 函館大学 (北海道)
- ⑤ 沼尾 雅之, 松澤裕史, 松尾総一郎: 集約バスケットからの相関関係マイニング, 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM2011), 2011 年 2 月 27 日, ラフォーレ修善寺 (静岡)
- ⑥ 鈴木克弥, 沼尾 雅之: 実世界トレーサビリティの為にクエリ処理システムの提案, 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM2011), 2011 年 2 月 27 日, ラフォーレ修善寺 (静岡)
- ⑦ 松尾総一郎, 沼尾 雅之: プロセス産業に向けた「履歴テーブル」を用いた分析手法の提案, 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM2011), 2011 年 2 月 27 日, ラフォーレ修善寺 (静岡)
- ⑧ M. Numao and K. Suzuki: Event-Based Integration of RFID and Sensor Data for 5W1H Traceability, The 3rd UEC-SJTU International Symposium on Robot and Artificial Intelligence, June 8, 2010, Shanghai (China)
- ⑨ 鈴木克弥, 沼尾 雅之: センサネットワークにおける問い合わせ言語を用いたデータ収集効率化手法の提案, 第 23 回人工知能学会全国大会, 2009 年 6 月 17 日, サンポートホール高松 (香川)
- ⑩ S. Hido, H. Matsuzawa, F. Kitayama, M. Numao: Trace Mining from Distributed

Assembly Databases for Causal Analysis,
The 13th Pacific-Asia Conference on
Knowledge Discovery and Data Mining
(PAKDD2009) (Refereed), April 27-30,
2009, Bangkok(Thailand)

〔図書〕 (計 0 件)

〔産業財産権〕

○出願状況 (計 0 件)

〔その他〕

ホームページ等

<http://www.nm.cs.uec.ac.jp>

6. 研究組織

(1) 研究代表者

沼尾 雅之 (NUMAO MASAYUKI)

電気通信大学・大学院情報理工学研究科・
教授

研究者番号： 9 0 5 0 8 8 2 1

(2) 研究分担者

尾内 理紀夫 (ONAI RIKIO)

電気通信大学・大学院情報理工学研究科・
教授

研究者番号： 7 0 3 2 3 8 7 1

(3) 連携研究者

なし