

令和 3 年 6 月 21 日現在

機関番号：13903

研究種目：基盤研究(C) (一般)

研究期間：2018～2020

課題番号：18K11432

研究課題名(和文)無限構造を扱うシステムの検証に向けた仕様発見系を持つプログラム推論技術の研究

研究課題名(英文) Specification Mining and Reasoning about Programs for Verifying Systems with Infinite Structures

研究代表者

世木 博久 (Seki, Hirohisa)

名古屋工業大学・工学(系)研究科(研究院)・教授

研究者番号：90242908

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：本研究では、リアクティブ・システムなどの無限に継続する状態を扱えるソフトウェアの正当性を検証するための方法論の確立に向けて、次のような3つの研究成果を得た。(1) プログラムの実行ログから構成される系列データ集合から再現ルール形式の仕様を発見するアルゴリズムについて、その効率化のためにルール形成処理を並列化する処理方式を実装した。(2) 先行研究の再現ルール発見アルゴリズムを再検討し、それを改善した新たなアルゴリズムを提案して、有効性を実験的に確認した。(3) バグのあるプログラムについて、その実行時トレース・データからバグの位置特定のための相関ルールを導出するアルゴリズムの設計した。

研究成果の学術的意義や社会的意義

プログラムやシステムの検証に必要な要素として、証明すべき仕様を導出する機能と、システム設計や実装の誤り(バグ)を効率的に発見する機能について、その実現アルゴリズムを提案し、検証システムを構築するための知見を蓄積した。これらの結果は、実応用に向けたソフトウェアの検証システムの実現に寄与する意義がある。

研究成果の概要(英文)：The objective of this research is to develop a methodology for verifying software such as reactive systems. We use some techniques of reasoning about programs for finding system specifications to be verified, and localizing potential bugs in a given program. In particular, we have obtained the following three main results: (1) we have proposed a parallel version of our previous algorithm for mining specification formulas from a sequence database of execution logs. (2) We have also proposed a new algorithm for mining specification formulas, which is more efficient than the previous algorithms in the literature. (3) We have then proposed a new method for mining association rules for finding bug locations in a given buggy program.

研究分野：知能情報学

キーワード：プログラム推論 パターン発見 形式概念分析

1. 研究開始当初の背景

ソフトウェアの高信頼性は、安全・安心な社会を実現するための基盤として不可欠なものとなっている。複雑化・高度化する様々なシステムを制御し運用するために、ソフトウェア・システムもますます複雑化しており、その安全性や信頼性を保証することは喫緊の課題である。従って、ソフトウェアの高信頼性に向けて、その正当性を保証するソフトウェア設計の方法論が必要となる。従来方法として、ソフトウェアの妥当性確認(バリデーション)やテストによって、誤りやバグを発見する方法が用いられてきているが、その方法では正当性を保証できない。安全性や信頼性を向上させるためには、ソフトウェアの正当性を形式的検証によって行う形式的手法に基づいたソフトウェア設計の方法論が必要となる。それに加えて、検証に必要な仕様を導出する方法と、システム設計や実装の誤り(バグ)を効率的に発見する方法を含めた枠組みの構築が必要である。

ソフトウェアの検証方法として、安全性が社会的にも経済的にもとりわけ重要なミッション・クリティカルなシステムであるリアクティブ・システム (reactive systems) を扱える検証の枠組を目指して、無限項を扱う確定余論理プログラム(co-logic programs) [Gupta et al. 07] に着目した。我々は、先行研究で、それに対するプログラム変換の新しい枠組みを提案している [Seki 11, 12]。この方法は特別な帰納法スキーマを用いない点で Jaffar らの従来法より単純で、従って実応用への有効性を持つ可能性がある。これらの結果を基にして、検証システムを完結するためには、検証すべき適切な仕様を与えることと、プログラムに誤り(バグ)がある場合にはその箇所を正確に見つけることが課題となる。

検証システムの実現に向けての前提として、対象システムの正確な仕様が与えられていなければならない。しかるに、システムの仕様を正確に記述して適切に与えることは難しい問題であることは知られている。その理由としては、いろいろな要因がある。まず、もともと仕様がきちんと記述されて残されていないこともしばしばある。また、仕様の記述が不完全で正確に表現できていない、更には、システムも時間の経過とともに、修正や機能の追加等により変更されるが、それに伴う仕様の変更がされないままになっていて、正確に反映されていないということもある。このように、対象とするシステムの正確な仕様が必ずしも与えられていないことが問題である。

また、ソフトウェアの規模が大きくなるにつれ、開発の過程において誤り(バグ)が混入することは避けられない。大規模なソフトウェアの場合、全体を複数のモジュールに分けて開発し、そのモジュールごとに適切なテストを用意するというような方法も用いられるが、この方法ではシステムのサイズが増大するに従ってデバッグのコストも増大する。このように、デバッグはソフトウェア開発においてコストがかかる作業であって、このデバッグ作業の効率化は、高品質なソフトウェア開発に大きく寄与する。デバッグ作業の効率化、そのなかでもプログラム中のバグの位置を特定する作業の効果的な自動化方法が望まれる。

以上のように、実際的な検証システムを構築するためには、先行研究の結果である無限構造を扱う論理プログラム推論系に加えて、適切な仕様を与える仕様発見系とバグ発見系の構築が必要となる。

2. 研究の目的

以上の背景から、ソフトウェアの安全性や信頼性を保証するために、その正当性を形式的に検証するプログラム推論技術を用いたソフトウェア開発の方法論を確立することを目的とする。特に、ソフトウェアの正当性を検証する形式的手法として、本研究では論理プログラムを対象としたプログラム推論技術を基礎とする。また、プログラム推論などの計算論理や形式手法、ソフトウェア工学などの知見とともに、人工知能、特に機械学習やパターンマイニングの分野の方法の適用可能性を探り、その過程を通して新たな方法の開拓を目指す。

先行研究で、我々はリアクティブ・システムなどを自然に表現でき無限構造を扱える確定余論理プログラムに着目した。そして、それに対するプログラム変換の枠組みを提案して、モデル検査等の推論に利用できることを示している。この推論系を基盤にして、本研究では、正当性が保証されたソフトウェア開発のための枠組みを構築することを目指す。まず、システムの諸性質を検証する対象となる仕様式について、それを自動的に導出する方法について検討する。本研究では、その仕様をパターンマイニング技術により発見し、それをシステム設計や実装の誤り(バグ)

発見に利用するアプローチを目指す。また、仕様検証の結果、プログラムに誤り(バグ)があることが分かった場合に、プログラム中のバグの位置を特定する作業の効果的に行うために、位置特定のためのルールを自動的に導出する方法の実現に向けて、その課題を明らかにすることを目的とする。

3. 研究の方法

本研究では、ソフトウェアの正当性を検証するために、プログラム推論技術を基盤として用いる。検証の入力として、対象システムとその証明すべき性質(仕様式)が与えられる。更に、対象システムの動作は論理プログラムを使って表現され、また、仕様式は一階述語論理式、あるいは時間論理式を使って表現されるとする。この時、仕様式を余論理プログラムに対する等価変換の規則を繰り返し適用することにより変換して、最終的にその真/偽を判定する。この推論系を基にして検証システムを完結するためには、検証に必要な仕様を導出することと、システム設計や実装の誤り(バグ)を効率的に発見する方法を含めた枠組みの構築が課題となる。本研究では、特に以下の三つのタスクに焦点を当てて研究を行った。

(1) 仕様発見アルゴリズムの実装方式の検討

検証に必要な仕様を導出するために、本研究では、実行ログの系列集合を格納した系列データベースに対して仕様マイニングを行うアプローチに着目している。仕様マイニングでは、通常の系列マイニングの場合よりも、CPU 処理時間やメモリ消費量のコストが大きい。従って、アルゴリズム実装の効率化についてさまざまな方向から検討する必要がある。

先行研究で、我々は従来方式を改良した仕様マイニング・アルゴリズム LF-NR3 を提案している。そのアルゴリズムに内在する並列性に着目して、それを利用した並列アルゴリズムの設計する。また、アルゴリズムでは、実行ログの系列に対する射影演算を含む文字列処理を多用している。そのような処理を効率的に行うためのデータ構造等の設計を行い、その有効性を実験的に検証する。実験に用いるデータとしては、系列パターンマイニングの分野で用いられているベンチマーク・データを中心にして、新たに提案する仕様マイニング・アルゴリズムと従来方法の実験結果を蓄積し比較する。また、本手法をプログラミング言語(Java)で PC 上に実装する場合の方式についても検討する。

(2) 仕様マイニングのためのパターン発見アルゴリズムの設計

仕様マイニング(仕様発見)の効率化について、前項の並列化によるアプローチに加えて、マイニング・アルゴリズム自体の見直しを含めて再検討する。従来研究の中で最良のアルゴリズムについて詳細な分析をすすめ、それを基にして新たな仕様マイニング・アルゴリズムを設計する。また、その実装について実験によって動特性の評価を行う。

(3) バグ発見のためのパターン発見アルゴリズムの実装方式の検討

バグ発見系では、プログラム中にバグがある場合に、その位置特定を行う。そのために、プログラム実行時の通過行の情報からなるログと、その実行の成功/失敗を示すラベルの組からなるトレース・データを入力として与える。本研究では、形式概念分析の方法を用いて、バグの位置を特定するための相関ルールを導出するアルゴリズムの設計を行う。また、その有効性を実験的に評価する。

4. 研究成果

本研究の主な成果は次の三つである。

(1) 仕様発見アルゴリズムの並列化の実装

先行研究において我々が提案した仕様マイニング・アルゴリズム LF-NR3 は、Lo たちによる従来方式 NR3 を改良したものである。ここでマイニングの対象とする仕様式は 前件 pre 後件 post という形式の再現ルール(recurrent rule)で、pre, post はイベント系列である。

この逐次版アルゴリズム LF-NR3 に基づき、それに内在する処理の並列性に着目した。すなわち、前件 pre が一つ見つくと直ちに、それに対応する後件 post を持つようなルール集合を構成するタスクを生成し、それをタスク・プールに格納する方法を採用した。タスク・プール中のタスクは、利用可能なワーカー・スレッドがあれば、その一つが直ちに実行されるという形式で設計した。この結果、ルール前件 pre の生成ごとに、それに継続するルール生成処理の並列実行が可能になった。

従来の素朴な並列処理、すなわち、前件パターン集合生成処理と後件パターン集合生成処理を並列に(パイプライン的に)行う方法と異なり、提案する並列化方式では、前件パターン pre の生成時に、pre とその際に計算される pre による射影データベースのデータも、継続するルール生成タスクに引数として受け渡している。これにより、ルール生成処理の計算の効率化を実現し

た。

提案アルゴリズムを Java 言語により実装して、その処理時間やメモリ使用量等の動特性を評価した。逐次版アルゴリズムや素朴な並列化方法と比較して、最大で 4.5 倍程度の処理速度の向上し、良好な結果を示した。

(2) 新しい仕様発見アルゴリズムの提案

前項の並列化手法とともに、再現ルールのマイニング・アルゴリズム自体を再検討するために、LoらのBOB (Bidirectional pruning-Based recurrent rule mining algorithm)アルゴリズムについて詳細な分析をすすめた。BOBは従来研究の中で最良のアルゴリズムの一つとして知られていて、まず前件パターン集合と後件パターン集合をすべて生成するという双方向マイニングを行い、次に、それらをペアにしてルール集合を形成するという2段階で処理されている。これについて、以下のような結果を得た。

LoらのBOBについての検討に基づき、再現ルールの生成を効率化した仕様発見アルゴリズム iBiRM (Interleaved Bidirectional Recurrent Rule Miner) を提案した。このアルゴリズムでは、BOBにおける二段階のステップをインターリーブする構造をしていて、これにより、計算時に必要となる中間的データの格納を除去する効率化を図った。

再現ルールマイニング時に必要な系列データベースのスキャン回数を削減するために、新しいサポートカウント法を導入した。従来法では、ルールの全ての前件パターン pre と全ての後件パターン post に対して、ルール pre post のサポートや確信度をカウントするために、与えられた系列データベースのスキャンが必要となる。提案方法では、後件パターンで部分系列を共有するものについては、1回のスキャンでサポートをまとめてカウントするようにし、スキャン回数の削減を図った。

提案アルゴリズム iBiRM を Java 言語により実装し、代表的な系列データセットを用いて、その処理時間やメモリ使用量等の動特性を評価した。実験では、従来版アルゴリズム BOB と比較して、処理速度とメモリ使用量の両面において改善を示した。

(3) バグ発見のためのパターン発見アルゴリズムの設計

本研究のバグ発見系では、プログラム中にバグがある場合に、その位置を特定することを目的とする。そのために、プログラム実行時の通過行の情報からなるログと、その実行の成功/失敗を示すラベルの組からなるトレース・データを入力とする。本研究では、形式概念分析の方法を用いて、バグの位置を特定するための相関ルールを導出するアルゴリズムの設計を行った。また、その有効性を実験的に評価し、以下のような結果を得た。

この問題を形式概念分析(FCA)の枠組で扱うために、実行時の通過行の情報からなるログ・データを文脈(context)として考える。FCAにおける関連性(relevancy)の概念を利用して、失敗に関連するプログラムの通過行からなる特徴集合を抽出するアルゴリズムを提案した。

提案アルゴリズムでは、FCAにおける概念計算アルゴリズムであるCb0に、関連性の性質に基づいた新たな枝刈り方式を導入して、関連パターン探索空間の減少を図った。また、提案アルゴリズムをJava言語により実装して、いくつかの実験によりその有効性を実験的に評価し、従来法よりも優れた性能を示すことを確認した。

また、FCAにおける類似度計算の手法と機械学習における特徴選択法を応用した方法を、自然言語処理のテキスト文書の分類問題に適用して、その有効性についても検討した。

以上の結果から、システムの検証に必要な仕様を導出することと、システム設計や実装の誤り(バグ)を効率的に発見する方法を含めた枠組みの構築のための知見を蓄積した。これらの結果は、実応用に向けた検証システムの実現に貢献する意義がある。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 H. Seki, S. Toriyama	4. 巻 10
2. 論文標題 Using term similarity measures for classifying short document data	5. 発行年 2021年
3. 雑誌名 International Journal of Computational Intelligence Studies	6. 最初と最後の頁 181-197
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 S.-Y. Yoon, H. Seki	4. 巻 23
2. 論文標題 A Parallel Algorithm for Mining Non-Redundant Recurrent Rules from a Sequence Database	5. 発行年 2019年
3. 雑誌名 Journal of Advanced Computational Intelligence and Intelligent Informatics	6. 最初と最後の頁 956-961
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 SeungYong Yoon, H. Seki	4. 巻 7
2. 論文標題 Mining Non-Redundant Recurrent Rules from a Sequence Database	5. 発行年 2018年
3. 雑誌名 International Journal of Computational Intelligence Studies	6. 最初と最後の頁 253-269
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計8件（うち招待講演 0件／うち国際学会 5件）

1. 発表者名 Hirohisa Seki, Taiki Yamada
2. 発表標題 A CbO-based Algorithm for Mining Class Relevant Patterns
3. 学会等名 15th Int'l. Conf. on Concept Lattices and Their Applications (CLA 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 鳥山 修平, 世木 博久,
2. 発表標題 単語拡張によるテキスト分類精度の改善と評価
3. 学会等名 第18回情報学ワークショップ (WiNF2020)
4. 発表年 2020年

1. 発表者名 鳥山 修平, 世木 博久,
2. 発表標題 単語拡張によるテキスト分類精度の改善と評価
3. 学会等名 情報処理学会第83回全国大会
4. 発表年 2021年

1. 発表者名 H. Seki, S.-Y. Yoon
2. 発表標題 A New Algorithm for Mining Recurrent Rules from a Sequence Database
3. 学会等名 2019 IEEE Int'l. Conf. on Systems, Man and Cybernetics (SMC 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 H. Seki, S. Toriyama
2. 発表標題 On Term Similarity Measures for Short Text Classification
3. 学会等名 2019 IEEE Int'l. Workshop on Computational Intelligence and Applications (IWCIA2019) (国際学会)
4. 発表年 2019年

1. 発表者名 H. Seki , M. Nagao
2. 発表標題 Mining Correlated Association Rules from Multi-Relational Data using Interval Patterns
3. 学会等名 14th Int'l. Conf. on Concept Lattices and Their Applications (CLA 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 S.-Y. Yoon , H. Seki
2. 発表標題 Efficient Mining of Recurrent Rules from a Sequence Database Using Multi-Core Processors
3. 学会等名 Joint 10th Int ' I. Conf. on Soft Computing and Intelligent Systems and 19th Int ' I Symp. on Advanced Intelligent Systems (SCIS&ISIS 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 鳥山 修平 世木 博久
2. 発表標題 テキスト分類のための文書拡張法の評価
3. 学会等名 情報処理学会第81回全国大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------