

PMSD: システム・ダイナミクスとプロセスマイニングを用いた

データ駆動型シミュレーション ★

Mahsa Pourbafrani, Wil M. P. van der Aalst

RWTH アーヘン大学プロセス・データサイエンス講座 (ドイツ)

{mahsa.bafrani,wvdaalst}@pads.rwth-aachen.de

概要. プロセスマイニングは、プロセスの発見や適合性の確認にとどまらず、ボトルネック分析や組織のマイニングの技術も提供しています。しかし、これらの技術のほとんどは後ろ向きのもので、PMSD は、前向きなシミュレーション技術をサポートする Web アプリケーション・ツールです。PMSD は、イベント・データやプロセスマイニングの結果を、実行・検証可能なシミュレーションモデルに変換します。PMSD には、ログの変換、時間窓の選択、関連性の検出、インタラクティブなモデルの生成、システム・ダイナミクスの形でのモデルのシミュレーションと検証、すなわち、集約されたシミュレーションの手法が含まれます。各モジュールの結果はツール上で可視化され、より良い解釈が可能となります。

キーワード: プロセスマイニング、シミュレーション、システム・ダイナミクス、What-if 分析

1 はじめに

プロセスマイニングは、組織に蓄積されたイベント・データ、すなわちイベントログを利用して、組織に実用的な洞察を提供するものです[1]。そしてさまざまなツールが、プロセスの発見、パフォーマンス分析、ボトルネック分析、逸脱の検出に対応しています。しかし、後ろ向きのプロセスマイニング技術と前向きのプロセスマイニング技術の間には、依然としてギャップがあります。[2]で述べたように、従来の前向きな技術は、プロセス内のイベントをシミュレーションの基礎として使用します。これは、プロセスを詳細なレベルで模倣し、シミュレーションを行うことを目的としていました。しかし、[3]のような最新のシ

★ Deutsche Forschungsgemeinschaft (DFG, ドイツ研究財団)による、ドイツのエクセレンス戦略-EXC 2023 Internet of Production-プロジェクト ID: 390621612 に基づいています。また、研究を支援してくださった Alexander von Humboldt (AvH) Stiftung 氏にも感謝いたします。

ミュレーションツールでは、アクティビティの継続時間やアクティビティの流れなど、異なるレベルの詳細情報を取得してシミュレーションを行います。さらに、発見されたペトリネットをシミュレーションするために、pm4py ツール¹では、モンテカルロ法が使用されています。

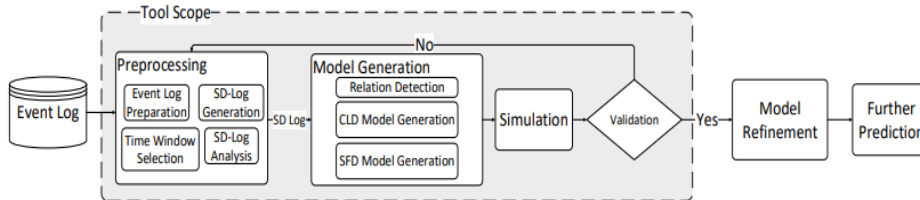


図 1. [4]で提案した、ビジネスプロセスのシナリオベースの予測を支援するための有効なモデルを設計するために、プロセスマイニングとシステム・ダイナミクスを併用するフレームワーク。本稿では、開発したツール、すなわち、ハイライトされたステップに焦点を当てる。

PMSD では、集約されたレベルのイベント・データからシミュレーションモデルを学習することができるという考え方を採用しています。従来のプロセスマイニングとシミュレーションの関係では、発見ステップで発見された記述モデルを用いて、分散イベントシミュレーション (DES) など、プロセスインスタンスのレベルでシミュレーションモデルを強化していました。今回発表したツールは、ビジネスプロセスのシミュレーション結果を集計レベルで生成し、シミュレーションに外部要因を追加するオプションを提供するという、我々のアプローチの成果です[4]。図 1 は、イベントログから始まり、シナリオベースのシミュレーションモデルで終わるアプローチの概要を示しています。ハイライトされている部分の手順は、本ツールによってサポートされています。図 2 に示すように、シミュレーションにイベントを考慮するのではなく、時間の異なるステップでプロセスから可能な変数を抽出します。

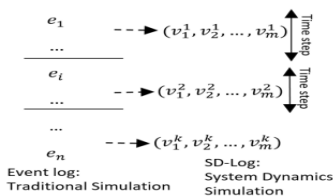


図 2. 従来のシミュレーションと PMSD の比較。時間ステップ(k)で可能な変数(m)を抽出します。

モデル生成モジュールは[6]で紹介されており、前処理ステップは[5]で紹介されています。イベントログは、時間の経過とともに一連の変数に変換され、これらの変数の値は、システムダイナミクスログ (SD-Log) を形成します。より安定した SD-Log を生成するために、我々は値の時系列解析を行います。SD-Log の時間経過に伴う変数間の関係は、システムダイナミクスモデルの作成に使用されます。これらは因果ループ図(CLD)と

¹ <http://pm4py.pads.rwth-aachen.de>

ストックフロー図(SFD)の両方に対応しています。システム・ダイナミクスでは、システムとその環境との関係をモデル化します[8]。CLD はこれらの概念的な関係を表現し、SFD はその基礎となる方程式をストック、フロー、変数の表記を用いてモデル化します。フローはストックの値を追加/削除し、変数はフローや他の変数に影響を与える/受けるものです。PMSD は、リソースの作業量とタスクの実行速度の間にある非線形関係など、ユーザーには見えない時間経過のプロセスを通して洞察を与えます。

2 機能性の説明

我々のアプローチでは、1日あたりの到着率や1日あたりの平均サービス時間など、可能性のあるプロセス変数を時系列で抽出します。新たに生成されたログ (SD-Log) は、シミュレーションの基礎となるものです。前処理ステップと、[5]で提案された時系列分析によって、フレームワーク内の最適なパラメータを抽出します。有効なシステムダイナミクスモデルを形成するためには、[6]で紹介されているように、生成されたプロセス変数の時間経過に伴うすべての関係、すなわち線形および非線形の相関関係を発見しなければなりません。プロセスを分析し、さらなる分析のためにプロセスの経時的な特徴 (プロセス変数) を集約して作成することが、このツールの主な目的です。

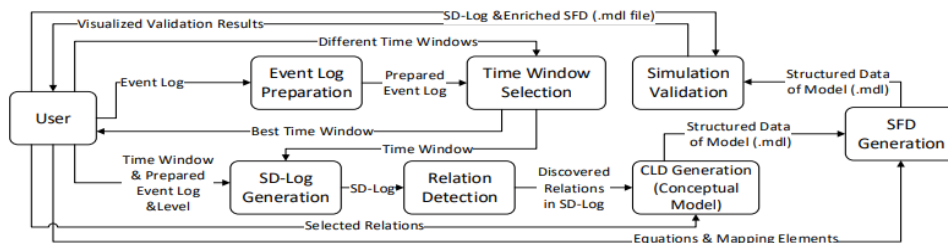


図 3. ユーザーとメインモジュール間のデータフローと、モジュール間のデータのバックグラウンドフローを含む PMSD のデータフロー図。

PMSD は、すべてのステップにおいて、ユーザーが出力にアクセスできるように設計されています。図 3 は、アプリケーションのデータ・フロー・ダイアグラムを描いたものです。各モジュールの入力と生成された出力、およびユーザーとのやり取りが示されています。プロセスのアクティブなステップや、選択された異なる時間ウィンドウのすべてのステップを含む、生成された SD-LOG が.csv 形式で取り込まれています。また、設計されたすべての CLD と SFD が.mdl 形式でユーザーのためにローカルに保存されます。このツールを

ローカルで実行するには、<http://127.0.0.1:5000> の URL を使用してブラウザからホームページにアクセスします。すべてのモジュールは異なるタブとして設計されており、視覚的にアクセスできます。PMSD は、Python と Flask の技術をベースにしたユーザーインターフェースを持つ、完全にインタラクティブなツールです。ステップの結果はグラフィカルに表示され、より簡単な解釈が可能です。PMSD には 8 つのタブがあり、それぞれのタブは他のモジュール/タブの異なる入力/出力で別々に実行することができます。現在、以下のコンポーネントが用意されています。

- *Event log transformation* では、イベントログの主な属性を示し、直接フォローするグラフを発見し、イベントログの情報を提示します。
- *Time window selection* は、シミュレーションデータを生成するための時間窓を選択するユーザーの好みの質を評価します。
- *Simulation log generator* は、変換されたイベントログと選択された時間窓を使用して、シミュレーション・データ (SD-Log) を生成します。SD-Log は、一般的なプロセス、組織、およびアクティビティの各側面とレベルに応じて生成されます。例えば、プロセスの一般的な側面の SD-Log には、プロセスの到着率、プロセスの平均サービス時間、およびその他の 1 日あたりの測定可能な変数が含まれます。
- *Relation detection* では、抽出された SD-Log 内の変数間に強い関係性があるかどうかを調査します。さらに、ユーザーは、時間の異なるステップにおける変数間の関係を探ることができます。
- *Detailed relations* では、関係の種類をさらに調査するために、SD-Log 内のすべての 2 つの変数間の既存の関係を提示します。
- *Interactive conceptual model generation* では、関係性検出モジュールで発見されたすべての強い関係性の中からユーザーが選択し、CLD、すなわちプロセス変数間の効果と関係性を作成するオプションが用意されています。このツールは、ツール内のグラフィカルモデルと、Vensim²などのほとんどのシステムダイナミクスツールで使用できる.mdl (テキストフォーマット) ファイルの両方を生成します。
- *Interactive stock-flow diagram* では、PMSD 内のグラフィカルな SFD と(.mdl) ファイルを生成します。その関係は CLD (前のステップ) から直接変換され、ユーザーはプロセス変数を SFD の要素にマッピングすることができます。
- *Simulation and validation* では、SD-Log の値を使って SFD モデルをシミュレーションし、SD-Log とシミュレーション結果の値とその分布のペアワイズ比較を使って結果を検証します。

3 ツールの成熟度

² www.vensim.com

我々が提案したプロセスマイニングの将来的なアプローチの評価結果は、このツールの様々なモジュールを使って表現されています。PMSD は [GitHub](#)³で公開されており、チュートリアルやスクリーンキャストも公開されています。また、PMSD は、Industry 4.0 に関連した Internet of Production のプロジェクトなど、いくつかの産業プロジェクトでも使用されています。[7]では、PMSD を生産ラインに使用した結果の一部が紹介されています。我々は CPN ツールで設計されたコールセンターのイベントログを例にして、同様の結果を示しています。

Stability Test of Time Window

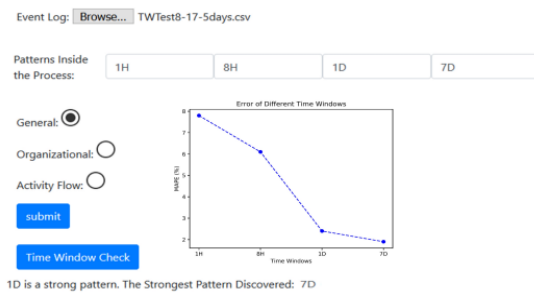
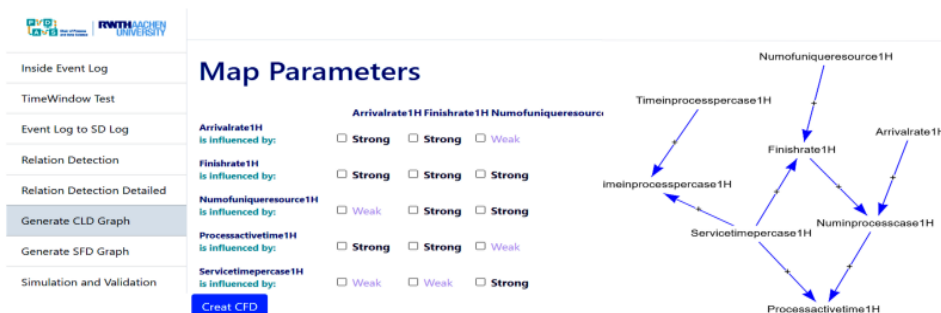


図 4. 各時間帯の学習モデルの誤差を示す安定性テスト。

我々は時間窓テストを用いて可能性のあるプロセス変数の時間的な値を抽出するために、異なる提案された時間窓を使用します。図 4 の結果は、ユーザーが選択した時間窓と、各時間窓に対する学習済みモデルの誤差を示しています。図 5 は、検出された変数間の強い関係を選択するためのユーザーインターフェースを示しています。最後に、生成された SFD と SD-log (いずれも自動生成) をアップロードすることで、自動シミュレーションが実行され、検証結果が検証モジュールに表示されます。検証結果には、実測値とシミュレーション値の比較や、選択した変数の分布などが含まれています。

4 おわりに

この論文では、ビジネスプロセスの文脈でシミュレーションを行うためのシステムダイナミクスモデルの設計を支援するための PMSD を紹介しました。PMSD を使用すること



³ <https://github.com/mbafrani/PMSD>

図 5. 検出された変数間の関係とその強さを示すコンセプト・モデリング・セクション。ユーザーは選択された関係の中から選択することができる。

で、我々は 1 時間単位や 1 日単位などの異なる集計レベルや、プロセス全体や組織的な側面などの異なる側面からプロセスを調べることができます。そして用意されたユーザーインターフェースとグラフィカルな出力により、結果の解釈が容易になります。PMSD を適用することで、インスタンスレベルでの根本的な効果や関係性を検出し、集約してモデル化することが可能です。このツールでは、モデルを直接シミュレーションして検証することができるほか、Vensim などのシミュレーション・ソフトウェアを使用して外部変数を追加することで、モデルをシミュレーションしたり、改良したりすることができます。

参考文献

1. van der Aalst, W.M.P.: Process Mining - Data Science in Action, Second Edition. Springer (2016)
2. van der Aalst, W.M.P.: Process Mining and Simulation: A Match Made in Heaven! In: Computer Simulation Conference. pp. 1–12. ACM Press (2018)
3. Camargo, M., Dumas, M., Rojas, O.G.: Simod: A tool for automated discovery of business process simulation models. In: Proceedings of Demonstration Track at BPM 2019. pp. 139–143 (2019)
4. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P.: Scenario-based prediction of business processes using system dynamics. In: On the Move to Meaningful Internet Systems: OTM 2019 Conferences - Confederated International Conferences: CoopIS, ODBASE, C&TC 2019, Rhodes, Greece, October 21-25, 2019, Proceedings. pp. 422–439 (2019). https://doi.org/10.1007/978-3-030-33246-4_27, https://doi.org/10.1007/978-3-030-33246-4_27
5. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P.: Semi-automated time-granularity detection for data-driven simulation using process mining and system dynamics. In: Conceptual Modeling - 39th International Conference, ER 2020, Vienna, Austria, November 3-6, 2020, Proceedings (2020)
6. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P.: Supporting automatic system dynamics model generation for simulation in the context of process mining. In: Business Information Systems - 23st International Conference, BIS 2020, Colorado Springs, USA, 8-10 June , 2020, Proceedings (2020)

7. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P.: Supporting decisions in production line processes by combining process mining and system dynamics. In: Proceedings of the 3rd International Conference on Intelligent Human Systems Integration. pp. 461–467 (2020). https://doi.org/10.1007/978-3-030-39512-4_72
8. Sterman, J.: System Dynamics: Systems Thinking and Modeling for a Complex World (2002)