A Framework for Implementing Process Mining and Robotic Process Automation in Organizations (Extended Abstract)

Najah Mary El-Gharib[1,*]

[1]*School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada*

Abstract

Process Mining (PM) is a technique that combines data science and process modeling to discover process models from system execution logs collecting activities and their timestamps, user identifiers, resources, and other parameters. PM is used to understand and improve processes. Robotic Process Automation (RPA) is an emerging technology that allows businesses to automate processes with software robots that replicate repetitive human tasks. RPA helps achieve accuracy, consistency, and efficiency in performing processes. Understanding processes is the key to automating them, but organizations often lack a deep understanding of their as-is processes and often rely on guess work to discover their automation opportunities. Using PM as a key enabler in discovering the processes that can be automated and in monitoring improvements is not an easy task and comes with several challenges. The main focus of this doctoral thesis is on introducing a new framework to help integrate the use of PM and RPA in an organization's context. The framework consists of several components including methods and tools.

Keywords

Process Mining, Robotic Process Automation, Integration

1. Introduction and Problem Definition

In the context of digital transformation, many organizations are automating their manual processes to improve performance, save costs, and minimize errors while executing these processes. However, there is currently a large amount of guess work in assessing the processes that can be automated and in monitoring their actual improvement. In the past five years, there has been a steep increase in the use of Robotic Process Automation (RPA) in organizations, together with related tool support [1]. RPA, which uses software robots to automate human tasks, has often been applied in the areas of administration and finance. RPA is frequently implemented in cases where high volumes of data are processed through repetitive tasks. The market of RPA solutions includes over 50 vendors who develop RPA tools that provide different functionalities to automate office tasks in an intelligent way [2].

Most organizations realized the need to automate their processes where it makes the most sense and where it positively impacts their business. The challenge here is not limited to how

---

to automate processes; it is also essential to understand which processes can benefit from automation. The problem addressed in this thesis is to tailor process mining techniques in order to discover candidate routines for automation using RPA, while taking into consideration the type of the processes and how organizations can implement the integration of the two technologies.

2. Research Goal and Research Questions

Having clear visibility into the organization's current processes before automating them is key to successful automation. Unfortunately, there is currently a large amount of guess work in assessing the processes that can be automated and in monitoring their actual improvements. Accordingly, this research aims to produce a framework that can leverage process mining techniques to discover candidate routines for automation. Its research questions are:

- **RQ1**: What are the challenges encountered by organizations that want to adopt process mining and robotic process automation technologies?
- **RQ2**: How can the synergy between process mining and robotic process automation accelerate the planning of automation projects and provide continuous monitoring?
- **RQ3**: How can process mining methods be tailored to discover the candidate routines that can be automated using robotic process automation?

3. Research Plan

3.1. Research Methodology

In order to conduct a good research project and study, an appropriate methodology is needed. Design Science Methodology in Information System [3], is selected for this PhD thesis since it aims to study the interaction between people, technology, and organizations. The framework proposed will be developed following this approach with several iterations at each step.

As the first step, in order to answer RQ1, we conducted a Systematic Literature Review (SLR) to assess the applicability of process mining algorithms in accelerating and improving the discovery of processes that can be automated and implementations using RPA [4]. The literature review focused on the methods used to record the events that occur at the level of the user interactions, and on the preprocessing methods used to discover candidate routines that can be automated. Additionally, the SLR covers the challenges that are encountered throughout the project life cycle. In order to further understand which processes are currently being automated within organizations and what challenges they face, how they are identifying processes that can be automated, we will be launching a survey to understand what processes organizations are currently automating using RPA and whether they are using any tools or techniques to discover the routines that can be automated. A subset of the survey respondents will also be involved in more detailed interviews. The purpose of conducting interviews is to help assess whether the proposed framework could potentially help organizations identify processes that can be automated using RPA. The outcomes and analysis of the survey results and the interviews will help in the definition and development of the proposed framework.

3.2. Proposed Framework

The framework focuses on how process mining can be integrated with RPA projects. Using the two solutions and integrating them into an organization's infrastructure is not a simple problem to solve. There are several components that play a role, and several challenges are expected at each phase. The framework will be structured from several components including 1) selected process, 2) datasets, 3) PM tool, 4) RPA tool, and 5) list of methods. For example, a data preprocessing method is needed, data restructuring can be another method needed, and several other methods will be identified as the problem is explored and investigated.

The first component is the process selected for process mining and automation. The datasets component represents the event logs that capture the selected process. This component requires selecting the systems that collect the datasets related to this particular process. The datasets component represents the event logs that capture the selected process. This component requires selecting the systems that collect the datasets related to this particular process. The desired level of granularity for the collected event logs is also required as this will determine the necessary preprocessing techniques that are needed. A full assessment is needed to identify the datasets related to a process.

PM tools are able to extract knowledge and generate process models from event logs collected in information systems. A list of requirements will be identified to help selecting the most appropriate tool to be used for analysis. These requirements will cover, for example, whether the tool is hosted on premise or on the cloud, its security features, and data source connectors. The conducted survey will help discover which PM tools are widely used within organizations. Similarly, an RPA tool is another component required to implement the software bots. Selecting the appropriate tool is another challenge and listing the criteria that can be used to select the tool is needed for this component. The features the tools provide will play a role in making decision on which tool to select. Additionally, this will help in assessing how well several PM tools can interact and integrate with RPA tools,to what extend they can be integrated together in an organization infrastructure, and what challenges are encountered.

Finally, the framework will include a list of methods that can be used throughout the project phases. Each method will have a list of inputs and will generate an output that will be used in another phase. The list of methods will include:

- Problem identification and formulation: based on the organization's objectives. For example, which processes can be automated and what is the success rate of this automation using RPA.
- Process domain selection: based organization's objectives and the problem that they are trying to solve; a process will be selected for analysis. Identifying the process that we are analyzing is the starting point.
- Requirements gathering: To understand the scope of the analysis and the project, this method includes identifying the information systems that collects the data associated with the selected process, as well as gathering the information about the stakeholders involved and the analysis questions that need to be answered.
- Datasets and traces collection: collecting datasets with the appropriate level of granularity is the starting point. Since we are looking to automation tasks performed by humans,

then the datasets and traces collected should be at the user interaction level with an application.

- PM and RPA tools selection: there are several tools in market to apply process mining and other set of tools to implement RPA solution, but selecting the right tools according to many factors within each organization is one of the main challenges.
- Preprocessing techniques for the dataset which are needed to identify the candidate routines that can be automated.
- Identifying whether a process can be automated based on a list of criteria such as the frequency of execution, the number of variants, and the number of exceptions.
- Identifying what process steps to automate. Several methods have been developed to identify candidate routines for automation building on exciting techniques [4, 5].
- Implementing new methods is another major component of this framework. Several methods are needed to record user interactions, converting the recorded data into event logs, preprocessing event logs to identify candidate routines for automation, implementing automation solutions. There are several methods that already exist, so I will be leveraging them in my solutions.
- Continuous monitoring for the robots after deployment. One of the main benefits of process mining is the continuous monitoring approach that it can help with, so the purpose is to leverage that after the automation deployment in order to help with monitoring the performance of the software bots that were implemented, in order to identify any errors, or unexpected behaviours.

3.3. Evaluation

In this research, my evaluation plan consists of a case study which is an observational evaluation method according to the Design Science Methodology [3]. The plan is to collaborate with an industry partner to conduct the case study on a selected process with real data. The case study could be followed by interviews to validate some of the results with different stakeholders. The purpose of this case study is to apply the methods and tools that are part of the framework explained previously. I might be doing several case studies to assess different types of processes to answer the research questions and attempt limited generalization. I am also considering applying the framework to one process that is executed in different organizations in order to compare whether the same methods can be used.

3.4. Technical Challenges

The challenges expected in this research are rooted in the recognized challenges of both process mining and robotic process automation. While conducting the systematic literature review, I extracted several challenges from the existing literature on process mining, RPA, and their intersection. Intersecting challenges include recording user interaction logs at an appropriate level of granularity, generating event logs from user interfaces, filtering noise, finding frequent patterns, extracting routines, segmenting event logs, and simplifying them. I expect other types of challenges to be identified from the survey and interviews.

4. Current Stage of this Research Project

An awareness of the thesis' research problem is developed after doing a literature review to examine the approaches and techniques where process mining techniques were used to understand the as-is processes that can be automated with software bots. Since the literature currently lacks a review that covers the intersection of process mining and robotic process automation, I conducted a systematic literature review that mainly focuses on the methods to record events that occur at a user interaction level, and on the preprocessing methods that are needed to discover routines that can be automated [6]. The second step is to conduct a survey in the North-American market to understand how the different organizations are currently implementing RPA, for example for which processes and with what success rates, in addition to process mining implementations, and how they identify if a process can be automated. Additionally, the survey will include questions such as organization size, industry, and process types. The survey will be followed by a series of interviews as explained above. After I finalized my research plan and proposal with the timeline, I will start the development of the framework components.

Acknowledgment

References

[1] W. M. Van der Aalst, M. Bichler, A. Heinzl, Robotic process automation, Business & Information Systems Engineering 60 (2018) 269–272.

[2] C. Dilmegani, Top 53 RPA tools / vendors & their features, 2022. https://research.aimultiple.com/rpa-tools/.

[3] K. Peffers, M. A. Rothenberger, W. L. K. Jr. (Eds.), DESRIST 2012, volume 7286 of *LNCS*, Springer, 2012. URL: https://doi.org/10.1007/978-3-642-29863-9. doi:10.1007/978-3-642-29863-9.

[4] A. Jimenez-Ramirez, H. A. Reijers, I. Barba, C. Del Valle, A method to improve the early stages of the robotic process automation lifecycle, in: CAiSE 2019, Springer, 2019, pp. 446–461.

[5] V. Leno, A. Augusto, M. Dumas, M. L. Rosa, F. M. Maggi, A. Polyvyanyy, Identifying candidate routines for robotic process automation from unsegmented UI logs, in: ICPM 2020, IEEE, 2020, pp. 153–160.

[6] N. M. El-Gharib, D. Amyot, A review of data-driven robotic process automation exploiting process mining, 2022. URL: https://arxiv.org/abs/2204.00751. doi:10.48550/ARXIV.2204.00751.