

Homework #1 (Scientific Article Summary)

Sep 2, 2022

ID : 2018204042 Name : 배홍섭Article Title : Robotic process MiningAuthors: Leno et al, BISE, 2021**1. Abstract**

Abstract 로봇 프로세스 자동화(RPA)는 웹 및 데스크탑 애플리케이션을 활용한, 조직에서 반복적인 사무 작업을 자동화할 수 있게 해주는 새로운 기술입니다. RPM의 핵심 아이디어는 사용자 상호작용(UI) 로그라고도 하는 작업자와 웹 및 데스크탑 애플리케이션 간의 상호작용 로그에서 자동화가 가능한 반복적인 루틴을 발견할 수 있다는 것입니다. 본 논문에서는 RPM을 뒷받침하는 일련의 기본 개념을 정의하고, RPM 도구가 UI 로그에서 RPA 스크립트를 생성할 수 있도록 하는 파이프라인을 제시합니다.

2. Introduction

RPM(Robotic Process Mining)이 나오게 된 background:

	A	B	C
1	Name	Surname	Country of residence
2	John	Doe	Australia
3	Albert	Rauf	Germany
4	Steven	Richards	Australia
5	Gerard	Dubois	France
6	Audrey	Backer	USA
7	Carl	Gustafsson	Sweden
8	Sarah	Johnson	Australia
9	Andrea	Bolzano	Italy
10	Hannah	Dietmeier	Germany
11	Igor	Honchar	Ukraine
12	Oliver	Dunkan	Ireland
13	Terry	Lee	Australia
14	Volodymyr	Leno	Ukraine
15	William	Macdonald	Canada
16	Jorge	Canales	Spain
17	Thomas	Taylor	Australia
18	Jack	Brown	Australia
19	Christina	Esposito	Italy
20	Amelia	Wilson	Australia

New Record

First Name

Last Name

Country of residence

☐ International Student

SAVE

(a) Student records spreadsheet (b) New Record creation form

다음과 같이 스프레드시트에 있는 정보를 하나씩 전송해야 될 경우 반복적인 사무 작업으로, RPA 도구를 사용하여 자동화할 수 있다.

Question:

- 수많은 routine 중에 자동화 대상을 어떻게 결정하나?
- 경험적 조사(근로자 인터뷰 및 관찰 등)는 시간이 너무 많이 소요됨.

따라서, 자동화를 위한 후보 루틴을 identify 하는데 RPA를 도구 사용

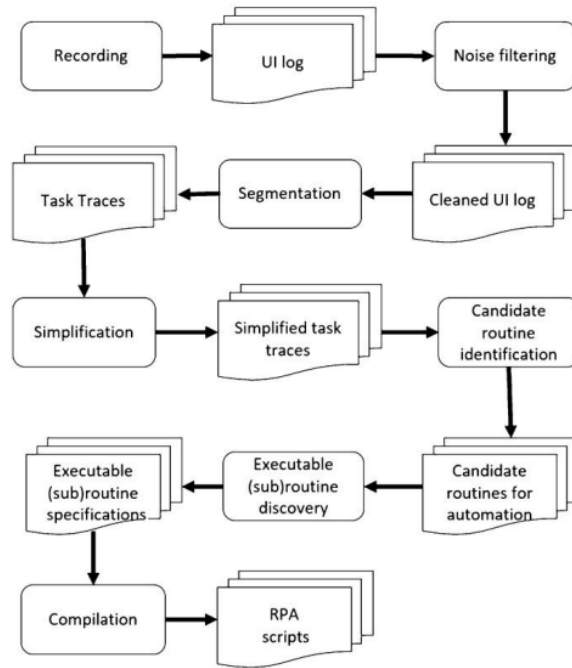


그림 3 RPM 파이프라인

-> RPM pipeline

RPM은 크게 3가지 단계로 구분

- 1) UI log 수집&전처리(recording, noise filtering, segmentation, simplification)
- 2) 후보 routine identification
- 3) Executable routine discovery

3. Method(pipeline 각 단계에서의 challenge와 solution or 사용된 전략)

1) Recording 단계

어떤 작업을 기록해야할지 identify 하는 것이 가장 중요. 기존 도구들은 너무 low level 정보만을 기록하였기 때문에 RPA tool을 사용, 그리고 high-level로 확장성을 위해 OCR 기술과 결합하여 사용할 수도 있다.

2) Noise filtering 단계

Noise가 routine의 시작 or 끝에 위치한 경우 noise 필터링 되지 않을 수 있다. 따라서 data object의 값을 고려하여 동일한 attribute를 공유하지 않은 데이터는 noise로 필터링 한다.

3) Segmentation 단계

- UI log와 Transactional data와 combine 하여 segmentation 수행
- Web session mining approach를 사용하여 segmentation 수행
- Use time 기반 휴리스틱 방식으로 segmentation 수행

4) Simplification 단계

이벤트가 task에 필수라도 중복될 수 있는 가능성이 존재함.

Sequential pattern mining 기술을 사용하여 이벤트와 이상치를 구분.

이때 거의 발생하지 않는 이벤트가 가끔 수행되면 이상치로 잘못 분류될 수 있음. 따라서 semantic filtering(의미론적 필터링)을 사용

5) Candidate routines identification 단계

- 첫번째 substep: routine을 나타내는 반복적인 patterns 식별 및 추출

Challenge 1: routine을 실행하는 동안 사용자가 routine의 일부가 아닌 다른 작업을 할 수도 있음(이메일 답장, 카카오톡 답장 등)

Solution: Liao and Chen 2013에서 연구된 gapped patterns과 함께 작동하는 sequential pattern mining techniques 사용.

Challenge 2: routine을 구성하는 작업의 순서가 임의로 변경될 수 있다.

Solution: 추상화 메커니즘(Bialy et al. 2019)를 사용.

- 두번째 substep: 자동화가 가능한 routine 식별

Challenge: 주요 기준은 작업의 실행 빈도. But 자동화가 가능하다는 보장이 없음.

Solution: 따라서 notion of determinism을 사용하여 효율적인 알고리즘을 설계하고 routine을 RPA에 적합하게 만드는 요소를 공식화하여 자동화 가능한 routine을 식별하는 방법 사용

6) Executable routine discovery 단계

Challenge: Routine sets가 주어졌을 때, 동일한 효과를 내는 다양한 routine이 존재할 수 있음.

Solution: 동일한 효과를 내는 모든 routine의 quality를 측정하여 최적의 routine을 식별

7) Compilation 단계(최종 컴파일 단계)

Routine 진행 중 application elements를 올바르게 인식해야함. 이때 OCR과 같은 기술이 사용될 수 있음.

4. Results(결과)

제안된 RPM 파이프라인은 RPA 봇이 end-to-end 방식으로 실행할 수 있는 루틴의 발견에 중점을 두고 있지만, 해당 assumption은 제한적이다. 실제로 루틴은 특정 하위 집합에 대해 자동화될 수 있지만 모든 경우에 대해 자동화되지는 못함(즉, 자동화는 부분적으로만 달성할 수 있음).

즉 주요 목표는 특정 subset에서만 실행가능한 routine을 발견하는 것이 아니라, 전체적으로 deterministic한 routine을 발견하는 것이다.

5. Discussion(토의)

- 본 논문이 가지는 가치

1) 본 논문은 RPA tool을 사용하여 자동화할 수 있는 routine을 식별하기 위해 IT 시스템과의 자세한 사용자 interaction UI 로그를 분석할 수 있는 새로운 클래스의 프로세스 마이닝 도구, 즉 RPM 도구에 대한 비전을 제시함.

2) 또한 이 비전을 구체화하기 위해 파이프라인을 분해하고, 각 단계에서의 method와 challenge를 제시함.

- 추가적인 challenges

1) 주어진 루틴을 자동화할 수 있는 조건을 찾는 아이디어를 아직 찾지 못했음.

2) RPM context에 적용가능한 data transformation을 만들기 위한 data transformations를 자동 발견할 수 있는 분야의 새로운 기술이 필요하다.