

RFID 추적을 위한 시간 기반 프로세스 마이닝

Time-based Process Mining for RFID Tracking

마황현(Hwang-Hyeon Ma)*, 오규협(KyuHyup Oh)*, 정재윤(Jae-Yoon Jung)**
{hyeon, k8383, jyjung}@hku.ac.kr

초록

기존의 프로세스 마이닝 기법들은 이벤트 발생 유무에 따라 비즈니스 프로세스 실행 과정을 도출하는 연구들이 주를 이룬다. 그러나 RFID 데이터와 같이 물리적 공간에서 사람이나 물체의 이동을 분석하는 경우에는 실행 경로보다 각 지점에서 체류한 시간이 더 의미있는 경우가 많다. 본 연구에서는 RFID 데이터를 대상으로 한 시간 기반 프로세스 마이닝 기법을 제시한다. RFID 데이터를 이용하여 각 지점에서의 체류 시간을 수집하고, 체류 지점의 시간에 따라 이동 경로들을 군집화하여 각 군집별로 프로세스 모델을 도출한다. 또한, 프로세스 모델은 사람 또는 자산의 위치별 체류 시간을 표현하여, 이동 경로뿐만 아니라 체류 시간을 표현할 수 있도록 설계하였다. 경로 분석에 치중한 기존의 프로세스 마이닝 연구들에 비하여, RFID 데이터와 같이 물리적 이동 경로를 대상으로 시간 기반 프로세스 마이닝을 제시하고 해결하였다.

1. 서론

RFID(Radio Frequency IDentification)는 사물을 식별할 수 있는 태그를 상품에 부착하고, 무선주파수를 이용하여 해석함으로써 상품과 주변 환경을 인식할 수 있는 기술이다. 최근 비즈니스 프로세스의 효율을 향상시키기 위한 기업의 RFID 도입이 확산됨에 따라 RFID를 사용한 비즈니스 프로세스 분석의 필요성이 점차 증가할 것이라 예상된다.

본 연구에서는 RFID 데이터를 이용한 시

간 기반 프로세스 마이닝 기법을 제시하였다. RFID 데이터를 이용하여 각 지점에서 체류한 시간을 수집하고, 이를 군집화하여 각 군집별로 프로세스 모델을 발견한다. 발견된 프로세스 모델은 사람 또는 자산의 체류 시간을 예측할 수 있도록, 이동 경로에 대한 시간 기반 분석을 지원한다.

경로 분석에 중심으로 프로세스 모델을 추출하는 기존의 프로세스 마이닝 연구들에 비하여, RFID 데이터를 바탕으로 시간 기반 프로세스 마이닝을 연구하였다는 데 본 연구의 의의가 있다.

본 연구는 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2010-0015334)

* 경희대학교 산업경영공학과 석사과정

** 경희대학교 산업경영공학과 교수

2. 관련 연구

비즈니스 프로세스 분석의 한 분야로서 프로세스 마이닝은 비즈니스 프로세스 관리(Business Process Management), 공급 사슬 관리(Supply Chain Management), 전사적 자원 관리(Enterprise Resources Planning) 등과 같은 다양한 정보시스템에 기록된 이벤트 로그를 분석하여 프로세스 모델을 추출하거나, 프로세스를 분석하는 연구이다. 프로세스 마이닝은 통계, 데이터 마이닝, 인공지능, 소셜 네트워크 등의 기법을 이용하여 핵심 프로세스 도출, 업무 성과 분석, 조직 분석 등 다양한 영역에 활용되고 있다[3,7].

RFID 프로세스 마이닝은 RFID 태그를 장착한 대상이 이동하는 과정에서 기록된 RFID 로그를 바탕으로 실제로 진행되는 비즈니스 프로세스 모델을 발견하거나, 비즈니스 성과를 분석하여 프로세스의 개선하고 향상시키기 위한 프로세스 마이닝의 적용 분야이다[2,5,6,8]. 기존의 RFID 프로세스 마이닝 연구는 RFID의 위치 정보로부터 비즈니스 프로세스 모델 추출하고 자원의 이동을 분석하는 등 경로 분석을 대상으로 연구가 진행된바 있다[5,6,8]

반면에 논문에서는 RFID 데이터의 시간 정보를 토대로 프로세스를 분석하는 방법을 제시한다. 먼저 RFID 데이터로부터 각 지점의 체류시간에 대한 정보를 수집하고, 프로세스 모델을 추출하기 위해 군집화 알고리즘 사용한다. 또한 발견된 각 군집 별로 프로세스 모델 도출하고 이동 경로에 대한 분석을 수행한다.

3. 시간 기반 프로세스마이닝

본 연구에서 제안하는 시간 기반 프로세스 마이닝은 <그림 1>과 같이 세 가지 단계로 수행된다. 먼저, 시스템에서 로그를 수집하여 시간 정보 추출하고, 계층적 군집화 알고리즘을 사용하여 프로세스 군집을 생성한 후, 마지막으로 각 군집들의 프로세스 모델을 도출한다.

3.1 시스템 로그 수집 및 전처리

시간 기반 프로세스 마이닝의 첫 번째 단계는 RFID 데이터의 전처리 과정이다. 전처리 과정은 시스템에 기록된 RFID 이벤트에서 위치, 시간, 태그 아이디를 추출한다. 이 과정에서는 태그를 부착한 대상이 RFID 리더가 설치된 지역에 나타난 시간과 사라진 시간의 차이를 구함으로써 체류 시간 정보를 추출한다. 이러한 과정을 통하여 로그에 기록된 데이터 중 불필요한 이벤트는 삭제하고, 시간 데이터만 추출하여 저장한다.



<그림 1> 프로세스 마이닝 과정

3.2 프로세스 군집화

이 단계에서는 RFID 태그 별로 결합한 위치별 체류 시간을 이용하여, 유사한 이동 경로 및 체류시간을 갖는 사람 또는 물체의 군집들을 생성한다. 본 연구에서는 계층적 군집화 알고리즘 중 하나인 CobWeb 알고리즘을 사용하였다 (<그림 2> 참조) [4].

```

FUNCTION COBWEB (Object, Root < of a classification tree>)
1) Update counts of the Root
2) IF Root is a leaf
    THEN Return the expanded leaf to accommodate the new object
    ELSE Find that child of Root that best hosts Object and
        perform one of the following
        a) Consider creating a new class and do so if appropriate
        b) Consider node merging and do so if appropriate and
            call COBWEB (Object, Merged node)
        c) Consider node splitting and do so if appropriate and
            call COBWEB (Object, Root)
        d) IF none of the above (a, b, or e) were performed
            THEN call COBWEB (Object, Best child of Root)

```

<그림 2> RFID 로그 군집화에 사용된 CobWeb 알고리즘 유사코드

CobWeb 알고리즘은 입력데이터를 하나씩 받아 유사한 데이터들을 클러스터링하는 방식으로 계층적인 구조, 하향식 분류, 자율 학습 등의 특징을 가진다[1,4].

다양한 알고리즘 중에서 CobWeb 알고리즘을 선택하여 군집화를 수행한 이유는 대상 데이터가 추가로 입력될 때마다 군집들이 업데이트하기 수월하기 때문이다. 즉, RFID 리더로부터 대상의 진행 경로가 하나씩 추가될 수 있는 RFID 기반 비즈니스 프로세스의 특징에 부합하기 때문이다.

3.3 프로세스 모델 도출

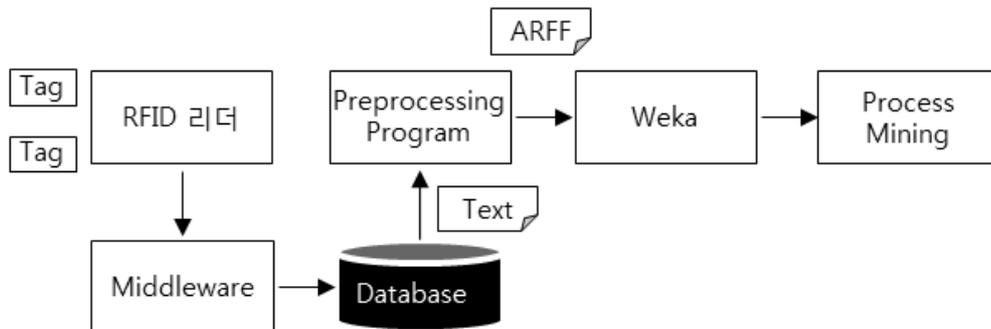
마지막 단계에서는 각 군집 별로 프로세스 모델을 도출한다. 시간 기반의 데이터를

통해 도출된 군집은 이동 위치뿐만 아니라 체류 시간 측면에서도 유사한 패턴을 가진다. 따라서 발견된 프로세스 모델은 위치를 기반으로 도출된 프로세스 모델과는 다른 형태를 가질 수 있다.

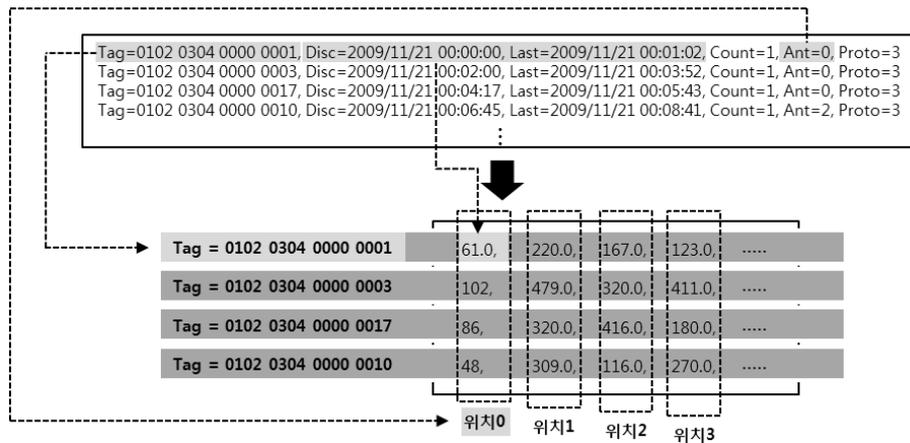
4. 시스템 구현

이 장에서는 3장에서 제안한 시간 기반의 프로세스 마이닝을 구현한 시스템을 설명한다.

<그림 3>는 개발된 시스템 구조를 보여준다. RFID 리더에 읽힌 태그 정보와 리딩 시간은 미들웨어 거쳐 데이터베이스에 저장된다. 그 후 RFID 로그를 분석에 적합한



<그림 3> RFID 시간 기반 프로세스 마이닝 시스템 작동 과정



<그림 4> RFID 추적을 위한 데이터 전처리 과정

형태로 변환하는 전처리 과정은 별도로 구현된 프로그램을 통해 수행된다.

<그림 4>은 전처리 과정에서 데이터가 변화하는 과정을 보여준다. 먼저 태그가 리더에 읽힌 시간과 리더에서 벗어난 시간의 차를 초단위로 계산한다. 계산된 값은 태그 아이디와 위치 데이터에 의해 저장할 자리가 정해진다. 즉, Disc와 Last의 시간차를 계산하여 체류시간을 구하고, Tag ID는 저장될 라인을, Ant는 컬럼의 위치를 정하는데 사용된다.

본 논문에서는 군집화를 처리하기 위하여 오픈소스 데이터마이닝 라이브러리인 Weka를 사용하였다[10]. Weka는 데이터마이닝의 다양한 알고리즘들을 자바로 구현한 오픈 소스 소프트웨어이다.

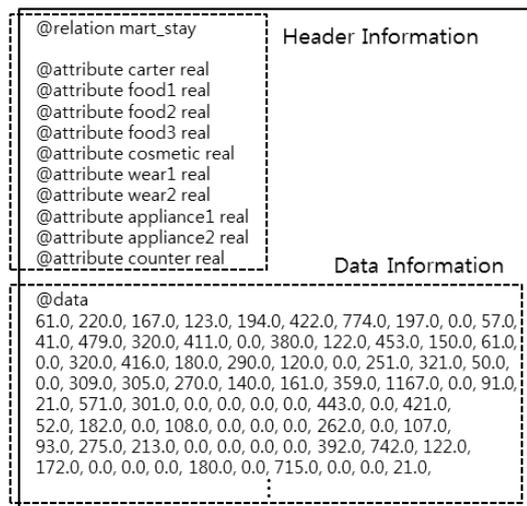
전처리가 끝난 RFID 체류시간 데이터는 Weka에서 사용하는 입력파일 형식인 ARFF로 변환된다. ARFF 파일은 인스턴스에 대한 속성 정보를 가지는 Header 부분과 실제 인스턴스를 저장하는 Data 부분으로 구분된다[10]. ARFF 형식으로 변환하는 과정은 다음과 같다. 먼저 전처리 과정을 마친 데이터를 ARFF 파일의 Data 부분으로 사용하고, 위치 정보를 포함하는 Header 부분을 추가하여 완성한다. <그림

5>는 완성된 ARFF 파일을 보여준다. ARFF 형식으로 변환된 데이터는 Weka에 입력되어 CobWeb 알고리즘을 군집화를 수행한다.

각 군집의 프로세스 인스턴스로부터 프로세스 모델을 추출하는 과정은 ProM 시스템을 사용하였다. ProM은 프로세스를 분석할 수 있는 다양한 분석환경을 제공하는 오픈소스 프로그램이다[9].

구현된 프로세스 마이닝 시스템은 다음장에서 예제를 이용한 실험 결과와 함께 상세히 설명한다.

ARFF format



<그림 5> 변환이 완료된 ARFF 파일

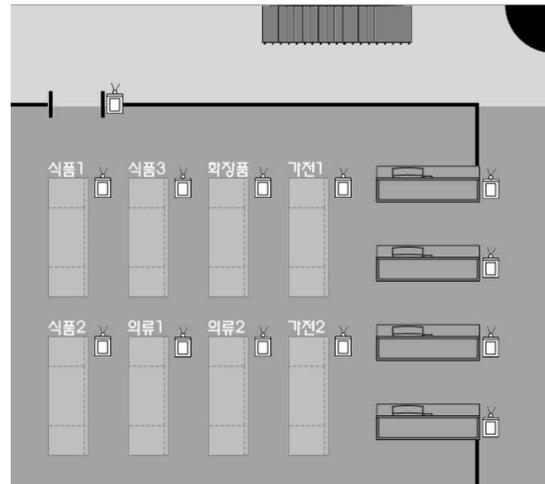
5. 적용 예제

본 장에서는 4장에서 제안한 시간 기반의 RFID 프로세스 마이닝의 수행 과정을 설명하기 위하여 예제를 제시하고 시간 기반 프로세스 마이닝 결과를 설명한다.

5.1 RFID 예제 설명

본 예제는 <그림 6>와 같이 RFID를 장착한 스마트 카트로 운영되는 대형 마트를 대상으로 설계하였다. 마트에 들어온 고객은 스마트 카트를 이용하고 각 판매대마다 RFID 리더기가 설치되어 있다고 가정하고 고객의 이동경로 및 시간을 바탕으로 프로세스 마이닝을 수행하였다.

예제를 통해 시스템 수행 단계를 살펴보면, 먼저 각 코너에 설치된 리더에서 카트에 장착된 태그정보를 인식하고 마들웨어로 전송한다. 전송된 태그 정보는 데이터베이스에 저장되고, 축적된 태그 정보는 전처리 과정을 통해 시간정보가 추출되고 ARFF 형식으로 변형된다. 최종적으로 Weka에서 CobWeb 알고리즘을 이용하여 군집화가 수행되고 각 군집 별로 프로세스 모델을 발견한다.

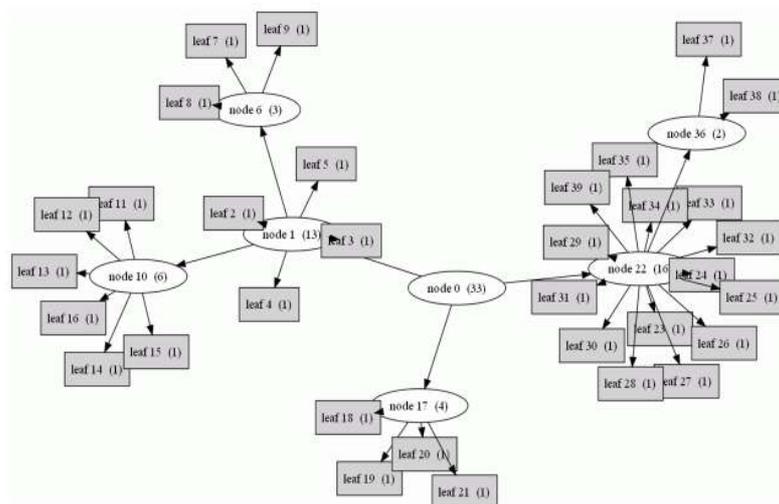


<그림 6> RFID가 설치된 대형마트 예제

5.2 분석 결과

5.1절에서 설명한 예제는 Weka의 CobWeb 알고리즘을 이용하여 <그림 7>와 같이 군집화되었다. 이 중 두 개의 군집을 선택하고 ProM의 Fuzzy Mining 알고리즘을 이용하여 프로세스 모델을 도출한 결과는 <그림 8>과 같다.

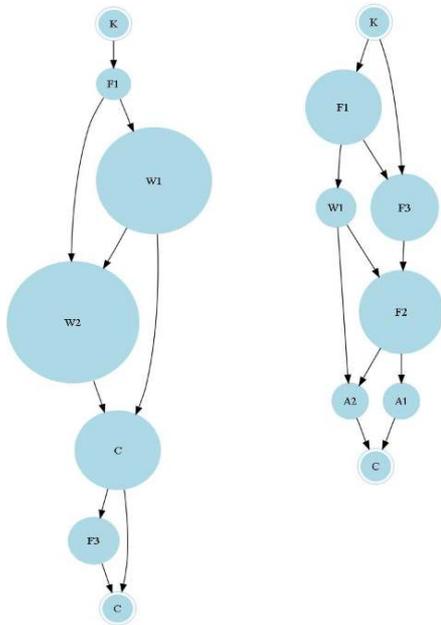
노드의 크기는 코너에서 체류한 시간을 나타내고, 화살표는 이동경로를 보여준다. 분석결과를 살펴보면, 모든 프로세스는 K에서 시작되어 C에서 종료된다. 분석결과 중 왼쪽의 프로세스는 주로 K(카트) → F1(식



<그림 7> CobWeb 알고리즘을 이용한 군집화 결과

품1) → W1(의류1) → W2(의류2) → C(화장품) → F3(식품3) → C(카운터)의 순서로 이동된다는 것을 알 수 있다. 하지만 F1과 F3의 노드 크기가 작기 때문에, F1과 F3 위치는 지나가는 과정이거나 고객에게 큰 관심이 없는 지점이라고 해석할 수 있다.

오른쪽 프로세스를 왼쪽과 같은 방식으로 분석하면 식품을 주로 구매하는 고객들의 프로세스에 나타난 W1, A1, A2 위치들이 고객의 관심을 끌어 이동한 것이 아니라는 것을 알 수 있다.



<그림 8> 각 군집의 프로세스 모델 비교

6. 결론

RFID의 도입은 체계적 비즈니스 관리를 위한 다양한 가능성을 보여주고 있다. 본 논문에서는 RFID로부터 수집된 자산의 흐름을 분석하기 위하여 체류 시간을 기반으로 프로세스를 분석하는 방법을 제시하였다.

본 연구의 방법론을 적용함으로써, 자산의 실제 이동 과정을 분석하고 비즈니스 프로세스의 유형 및 경과 시간을 분류하여 체

계적으로 관리할 수 있을 것으로 기대된다.

참고문헌

- [1] 백혜정, 박영택, “점진적 개념학습의 클러스터 응집도 개선”, 정보처리학회지, 제10권, 제3호, pp.297-304, 2003.
- [2] Bi, H.H. and Lin, D.K.J., “RFID-Enabled Discovery of Supply Networks”, IEEE Trans. on Engineering Management, Vol. 56, No. 1, pp.129-141, 2009.
- [3] de Medeiros, A.K.A, Weijters, A. J. M.M., and van der Aalst, W.M.P., “Genetic Process Mining: An Experimental Evaluation,” Data Mining and Knowledge Discovery, Vol. 14, No. 2. 2007
- [4] Fisher, D.H., "Knowledge acquisition via incremental conceptual clustering," Machine Learning, Vol. 2, pp.139-172, 1987.
- [5] Gerke, K., Claus, A., and Mending, J., “Process Mining of RFID-based Supply Chains,” In Proc. of IEEE Conf. on Commerce and Enterprise Computing, pp.285-292, 2009.
- [6] Hector, G., Han, J., and Li, X., “Mining Compress Commodity Workflows from Massive RFID Data Sets”, Proc. of Int’l Conf. on Information and Knowledge Management, pp.162-171, 2006.
- [7] Jansen-Vullers, M. H., van der Aalst, W.M.P., and Rosemann, M., “Mining Configurable Enterprise Information Systems”, Data and Knowledge Engineering, Vol. 56, No. 3, pp.195-244, 2006.
- [8] Oh, K. and Jung, J.-Y. "A Framework of Process Mining for RFID Event Analysis", In Proc. of Int’l Conf. on Industrial Engineering and Operations Management, pp.681-685, 2011.
- [9] ProM Homepage.
<http://prom.sourceforge.net>
- [10] Weka Homepage,

<http://www.cs.waikato.ac.nz/ml/weka/>