

# 워크플로우 모니터링 기반 전자상거래 프로세스 마이닝 알고리즘

홍형석<sup>o</sup> 김상배 배성용 김광훈 백수기  
경기대학교 전자계산학과  
{hshong, sbkim, sybae, kwang, skpaik}@kyonggi.ac.kr

## E-Commerce Process Mining Algorithm using Workflow Monitoring Services

Hyeong-Seok Hong<sup>o</sup> Sang-Bae Kim Sung-Yong Bae Kwang-Hoon Kim Su-Ki Paik  
Dept. of Computer Science, Kyonggi University

### 요약

워크플로우의 응용 분야 중에서 전자상거래의 B2B나 B2C에 적용하는 사례가 빈번히 발생하는 추세이다. 이에 워크플로우의 모니터링을 이용하여 전자상거래의 프로세스의 경향을 분석하여 기업에게는 고객의 분석, 고객에게는 상품의 선택에 있어서 정보를 제공할 수 있다. 이러한 일련의 과정을 워크플로우를 적용한 전자상거래의 프로세스 마이닝이라 할 수 있다. 프로세스 마이닝을 위해 내부 데이터 흐름 즉, 워크플로우의 관련데이터를 모니터링하여 제공해야 한다. 또한, 기존의 모니터링의 기능 중에 프로세스의 흐름을 분석하는 기능을 추가해야 한다. 본 논문에서는 프로세스 마이닝이 하는 알고리즘을 일반화 하여 전자상거래의 최대 관심사인 CRM을 구축하는 정보로 제공한다.

### 1. 서론

최근 워크플로우 기술은 사무 정보 시스템과 데이터베이스 시스템 분야에서 관심의 대상이 되고 있다. 워크플로우는 정상적으로 액티비티의 집합으로 이루어진 비즈니스의 절차를 기술한 것이다. 각각의 액티비티는 워크플로우 엔진의 제어를 통해 자동적으로 수행될 수 있다. 또한, 워크플로우 관리 시스템은 사전의 분석과 요약, 정보 중심의 작업과 액티비티의 자동화에 대한 도구의 집합이다. 최근 워크플로우 개발 환경의 동향을 살펴보면, 두 가지 현상으로 요약할 수 있다. 하나는 네트워크에서의 개인 컴퓨팅 환경이 강화되고, 워크플로우 응용 분야는 대규모화 되고, 복잡해 지고 있다. 더 나아가, 워크플로우 시스템의 기술은 앞에 제시한 동향에 맞추어 변화되고 있다. 즉, 일반적으로 워크플로우 시스템은 제어, 기술, 데이터 조작 기술에서 분산 구조를 가지며 개념적으로 내포된 워크플로우 모델의 구체화에 의해 설계되고, 물리적으로는 객체 지향 방법론에 의해 구현된다. 또한, 전적으로 웹 기반 환경 하에 동작을 한다. 최근에 와서 B2C(Business-To-Customer)/B2B(Business-To-Business)로 대변되는 전자상거래 및 전자시장(E-Market Place)의 활성화가 급속하게 확장됨에 따라 워크플로우를 기반으로 한 기업 내부 업무 흐름의 자동화를 의미하는 B2E(Business-To-Enterprise)의 구축을 더욱 가속화시키고 있으며, 워크플로우 기술 및 시스템은 결국 기업 또는 조직체내의 모든 업무 처리 절차(business procedure)들을 통합 관리하는 인프라구조(Infrastructure)로 인식되고 있다.

이러한 변화되는 환경 하에서 워크플로우의 중요한 요소 중에 하나인 워크플로우 모니터링 역시 변화가 요구된다. 기존의 워크플로우 모니터링의 기능은 단순히 프로세스의 상태정보나 감시정보만을 제공하는 것이었다. 하지만 전자상거래에서 모니터링의 역할은 프로세스를 분석하여 기업이나 고객에게 또 다른 정보를 제공할 수 있어야 한다. 이에 본 논문은 워크플로우의 모니터링 기능을 확장하여 워크플로우를 정의하는 부분에서만 제공하였던 프로세스에 대한 분석을 기존의 모니터링 정보를 바탕으로 하는 프로세스의 분석정보 다시 말해서 프로세스 마이닝의 기법에 대하여 기술 하고자 한다. 이를 위해서는 추가되는 모니터링 정보가 요구되고, 마이닝의 단계 및 시점을 정의 하여야 한다. 본 논문의 범위는 프로세스 마이닝의 시작 단계 즉 정의된 프로세스를 분석하기 위한 정보로 변형하기 위한 알고리즘을 제시한다.

2장에서 워크플로우 프로세스 모니터링에서 추가로 수집해야 하는 정보를 제시하고, 3장에서는 모니터링 정보를 바탕으로 하는 프로세스 마이닝 기법에 대해 소개 한다. 4장에서는 마이닝을 하기 위한 프로세스의 변형 알고리즘을 예제를 통해 제시하고, 5장에서 결론을 맺는다.

### 2. 프로세스 마이닝을 위한 모니터링 데이터

일반적으로 워크플로우 모니터링이라 함은 워크플로우 시스템에서 일어나는 이벤트를 추적하고, 기록하

는 능력을 말한다. 워크플로우 모니터링 서비스는 시스템 수준의 모니터링부터 사용자 수준 혹은 응용 프로그램 수준의 모니터링 서비스까지 매우 다양하다. 시스템 수준의 전형적인 예를 들면 실행 요소들 즉, 분산 엔진등의 수행도나 적재량을 모니터 하는 것이다.

그리고, 워크플로우 인스턴스와 관련하여 각각의 액티비티의 상태를 보여주는 서비스는 전형적인 사용자 수준의 모니터링이라 한다. 그러나 앞장에서 언급한 바와 같이 전자상거래를 워크플로우에 적용하면, 모니터링의 기능이 단지 상태정보나 감시정보를 제공하는 것 뿐 아니라 새로운 기능이 추가 되어야 한다. 예를들어 고객의 주문에 대한 경향을 분석하는 것이다. 이는 궁극적으로 전자상거래에서의 CRM(Customer Relationship Management)와 연관을 갖는다. 따라서 워크플로우의 모니터링은 CRM의 구현을 도울 수 있다. 앞에서 제시한 기능은 단순한 기능이지만, 모니터링 감시정보를 바탕으로 프로세스에 대한 분석을 하여 기업에 CRM을 위한 중요한 마이닝 정보를 제공할 수 있다.

전자상거래 프로세스의 특징은 주문 프로세스 주류를 이루고 있기 때문에 프로세스의 관점에서는 어떠한 경로로 어떠한 상품을 주문하였는지가 주요 관심 대상이다. 따라서 모니터링의 감시 정보 중 필요한 것은 프로세스의 흐름과 데이터의 흐름 정보이다. 이 두 가지 정보를 바탕으로 마이닝을 하는데, 프로세스의 흐름을 외부데이터, 데이터의 흐름을 내부 데이터라 한다.

3.1 내부 데이터 수집

내부 데이터는 워크플로우의 데이터 흐름 정보를 말하는데, 이것은 워크플로우 내부에서 일어나는 관련데이터(Relevant Data)를 의미한다. 이 관련 데이터는 주로 고객의 정보와 상품의 정보를 나타낸다. 따라서 모니터링은 고객의 나이, 성별등의 정보와 상품정보를 별도로 저장하는 것이 필요하다. 또한, 저장소에 저장시 데이터를 그대로 저장하는 것이 아니라, 나이와 상품, 성별과 상품 등의 연관관계를 정의하여 저장해야 한다.

3.2 외부 데이터 수집

외부 데이터는 워크플로우의 프로세스 흐름 정보를 말하는데, 이것은 워크플로우 외부에서 일어나는 전이정보(Transition Information)를 의미한다. 이것을 이용하여 마이닝을 위해서는 고객이 어떠한 프로세스의 경로를 선택하였는지가 중요하고, 이것을 모니터링해야 한다. 이는 프로세스가 시작하기 전에 모든 경로에 대한 정보를 저장하여, 고객이 어떠한 경로를 통해 주문을 하였는지에 대해 정보를 수집하고 저장한다.

3. 프로세스 마이닝 기법

프로세스 마이닝이라 함은 워크플로우에서 발생하는 데이터를 바탕으로 프로세스에 관련한 새로운 데이터를 발생하는 것이다. 이를 위해 앞장에서 정의된 데이터를 바탕으로 알고리즘을 통해 새로운 데이터를 생성하고, 워크플로우 로그 정보와 비교하여 프로세스에 관련된 새로운 데이터를 생성한다. 이를 구체적으로 단계별로 나열하면, 다음과 같다.

1단계: 정의된 프로세스를 필요한 데이터로 구분한다.

- 2단계 : 구분된 데이터를 알고리즘을 통해 새로운 데이터를 생성하고, 저장한다.
- 3단계 : 프로세스의 실행 중의 데이터를 모니터링 하여 저장한다.
- 4단계 : 알고리즘을 통해 생성된 데이터와 모니터링한 정보를 비교하여 저장한다.
- 5단계 : 저장된 비교 데이터를 CRM을 위한 정보로 변형시킨다.

1단계에서는 정의된 워크플로우를 경향을 찾기 위해 프로세스의 전이정보(Transition Information)를 나열하는 단계이다. 1단계에서 수집된 정보를 바탕으로 2단계에서는 프로세스의 모든 경로를 찾기 위한 알고리즘을 통해 모든 경로를 나열하고, 저장한다. 이 알고리즘에 대한 내용은 다음 장에서 자세히 설명한다. 3단계에서는 모니터링의 기본 기능인 프로세스의 상태 정보를 바탕으로 액티비티가 실행된 것과 어떠한 상품을 주문하였는지에 대해 모두 모니터링하여 로그정보를 저장한다. 4단계에서는 알고리즘을 통해 생성된 데이터와 모니터링한 정보를 비교하여 어떠한 경로를 통해 수행이 되었는지를 파악한다. 5단계에서는 4단계에서 저장된 정보를 어떠한 경로가 어떠한 상품으로 얼마만큼의 주문이 요청되었는지에 대해 분석한다. 이 정보는 CRM을 위한 정보를 쓰이게 된다.

4. 경로 분석을 위한 알고리즘

본 장에서는 앞장에서 제시한 단계 중에서 전이정보를 바탕으로 프로세스 경향을 찾고자 하는 경로를 나열하는 2단계에 대한 알고리즘을 예제를 통해 제시한다.

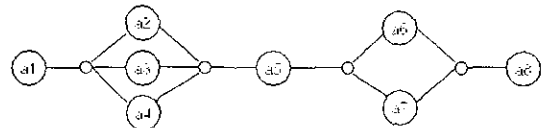


그림 1. 예제 프로세스

그림 2는 알고리즘을 적용하기 위한 가상의 예제이다. 표기는 ICN(Information Control Net)을 이용한다. a1부터 a8까지는 액티비티를 의미하고, 작은 원은 분기점과 집합점을 의미한다. 프로세스는 왼쪽에서 오른쪽으로 수행된다. 따라서 시작점은 a1이다.

위의 프로세스를 열거 하기 위해서 먼저 a1부터 검색을 시작한다.



그림 2. 초기 상태

알고리즘 1 : 처음 시작 액티비티를 트리의 노드로 구성한다.

액티비티가 시작점일 경우 그림 2와 같이 트리의 루트(root)로 구성하고 다음 액티비티를 검색한다. 다음 액티비티는 프로세스의 전이 정보(Transition Information)을 이용하여 검색한다. 다음 액티비티는 분기 액티비티이다.

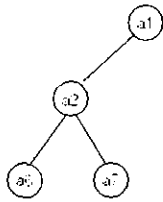


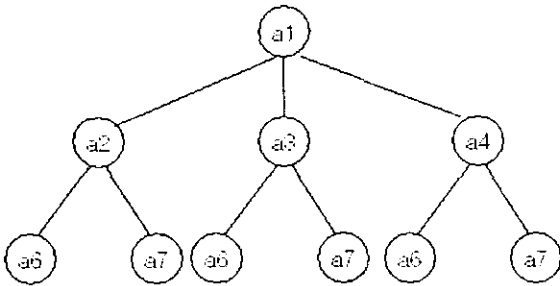
그림 3. 첫번째 분기점

알고리즘 2 : 분기 액티비티는 분기 수 만큼 수행을 하게 되는데, 먼저 분기 액티비티를 노드로 구성을 하고, 그 노드의 부모노드를 기존의 노드로 구성한다.

알고리즘 3 : 일반 액티비티나 집합 액티비티는 노드로 구성하지 않고, 다음 액티비티를 검색한다.

알고리즘 4 : 끝 액티비티는 현재의 트리를 반환한다.

분기점을 만나게 되면, 그림 3과 같이 a2를 구성하고, 부모노드로 a1을 설정한다. 또한 끝 액티비티가 검색될 때 까지 검색을 한다. 끝 노드를 만나게 되면, 지금까지 구성된 트리를 반환한다. 따라서, a6과 a7을 구성하고, a1의 다음 액티비티를 검색한다. 다음 액티비티는 a3인데, 하위 노드는 앞에 설명한 a2의 경우와 동일 하므로 같은 형태의 트리가 생성된다. 이렇게 트리를 구성하면 그림 4와 같다.



total path = The number of leaf nodes

- path1 : a1, a2, a6
- path2 : a1, a2, a7
- path3 : a1, a3, a6
- path4 : a1, a3, a7
- path3 : a1, a4, a6
- path4 : a1, a4, a7

그림 4. 알고리즘 적용 후 결과

이렇게 하여 트리를 구성하면, 프로세스에 대한 총 경로의 수가 결정된다. 경로의 수는 마지막 노드의 수이고, 그에 대한 경로를 루트노드에서 마지막 노드로 가는 경로를 나열한 것이다. 위의 예제에서는 총 경로의 개수는 6개이고, 그에 대한 경로를 그림 4에서 나타냈다. 위의 설명된 알고리즘을 일반화한 것은 표1과 같다. ProcessToTree 알고리즘은 재귀 호출 방식을 이용하여 처음부터 끝까지의 액티비티를 검색하여 트리로 구성 할 수 있다. 트리는 일반 트리로서 여러 자식노드를 가질 수 있는 트리로서 구성된다.

표1 경로 검색 알고리즘

```

ProcessToTree (Input Activity A, Node N)
begin
  if A is start Activity
    begin
      set A to N
      N is ProcessToTree(Next, N)
    else if A is normal Activity or join Activity
      begin
        N is ProcessToTree(Next, N)
      end
    else if A is split Activity
      begin
        for 1 to the number of split do
          begin
            set Next to N1
            parent of N1 is N
            N1 is ProcessToTree(Next of Next, N1)
            add N1 to children of N
          end
        end
      end
    else if A is end Activity
      begin return N end
    return N
  end

```

위와 같은 방법으로 프로세스의 경향을 분석하기 위한 데이터를 추출하고, 모니터링의 정보를 비교하게 되면, 고객의 경향을 파악할 수 있는 중요한 데이터로 변하게 된다. 이는 기업이 어떠한 고객이 어떠한 상품을 통해 어떠한 경로 통해 주문을 하였는지에 대한 정보를 파악할 수 있으므로 궁극적으로는 CRM의 구현을 이룰 수 있다.

5. 결론 및 향후 발전 방향

본 논문은 워크플로우 추세에 맞추어 변화하는 워크플로우 시스템에 적용하기 위한 모니터링의 추가적인 기능의 하나로 프로세스 마이닝 기법에 대해 소개하고, 마이닝을 위한 경향 분석 알고리즘을 제시하였다. 프로세스 마이닝을 위해서는 내부 데이터와 외부 데이터를 통해 가능해 진다. 이것은 궁극적으로 CRM을 구현하는 것과 일맥상통하는데, 전자상거래에서 워크플로우를 적용하는 장점 중에 하나로 그 역할을 담당할 수 있다.

앞으로 본 논문에서 제시한 알고리즘을 바탕으로 프로세스의 마이닝 단계를 적용하고, 엔진에서 지나야 할 데이터를 정의 할 것이다. 이는 워크플로우 관리 시스템 표준에서 엔진의 새로운 정보로 유용한 가치를 창출할 수 있다.

6. 참고문헌

- [1] Workflow Management Coalition Specification Document, " The Workflow Reference Model.", Version 1.1, November 1994
- [2] H S Hong, B S Lee, K H Kim, S K Paik, "A Web-based Transaction Workflow Monitoring System" WISE 2000 in Hong Kong Proceeding, IEEE PRESS
- [3] Rakesh Agrawal, Dimitrios Gunopulos "Mining Process Models From Workflow Logs", IBM Almaden Research Center, San Joes, CA, USA