

## 워크플로우 마이닝을 위한 제어경로 분석 알고리즘

박민재<sup>0</sup> 안형진 원재강 김광훈  
 경기대학교 전자계산학과 워크플로우기술연구실  
 {mjpark<sup>0</sup>, ctrl\_workflow, jkwon, kwang}@kyonggi.ac.kr

### Control-path Analysis Algorithm for Workflow Mining

Minjae Park<sup>0</sup> Hyungjin Ahn Jaekang Won Kwanghoon Kim  
 Dept. of Computer Science, Kyonggi University

#### 요 약

본 논문에서는 워크플로우 마이닝을 위한 제어경로 분석 알고리즘에 관하여 기술한다. ICN(Information Control Nets)은 워크플로우를 모델링하기 위한 방법이다. 본 논문은 ICN 기반의 워크플로우 모델 정보로부터 워크플로우의 제어경로를 파악하기 위한 제어경로 분석 알고리즘을 제안한다. 제어경로는 실제 워크케이스로 발생할 수 있는 워크플로우 모델의 유형을 의미하며, 제어경로 분석은 각 워크케이스의 유형을 분류해 내는 것을 의미한다. 분류된 워크케이스의 유형 정보는 워크플로우 로그와의 비교 분석을 통해 마이닝에 적용할 수 있다.

#### 1. 서 론

최근 몇 년 동안 국내외의 많은 기업들은 워크플로우 및 BPM을 도입하였고, 그로 인해, 비즈니스 프로세스의 가시화와 투명화, 비즈니스 프로세스 실행비용의 감소화, 비즈니스 프로세스의 가속화, 자동화, 생산성 향상, 비즈니스 프로세스의 복잡도 개선, 비즈니스 프로세스 실행 오류의 감소, 비즈니스 프로세스의 투명한 관리 및 외부 비즈니스와의 조정으로 기업 품질 자체를 향상시켜 고객 만족도를 증가 시켰으며, 변화에 대한 민첩한 대응력 및 유연성을 갖추었다. 하지만, 정의 시에 정의된 프로세스 정의정보와 워크플로우 제정 서비스(Workflow Enactment Service)의 실행 시 발생하는 프로세스 인스턴스와의 불일치가 일어난다는 점을 발견하게 되었다. 이는 워크플로우 모델링 도구를 통해 정의한 프로세스 모델이 실제로 일어나는 비즈니스 프로세스를 반영한 최적화 된 모델이 아니라는 것을 의미하며, 정의한 프로세스 모델은 분명히 개선이 필요하다는 것을 의미한다. 프로세스의 정의정보와 실행정보 사이의 불일치를 해소하고, 프로세스를 개선하기 위한 목적으로 워크플로우 웨어하우징, 워크플로우 마이닝 기술이 등장하였다.

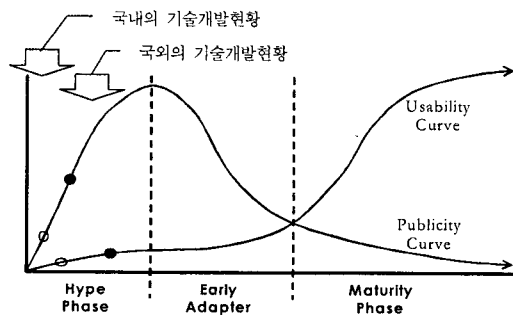


그림 1. 워크플로우 마이닝의 국내외의 개발 현황

그림 1은 Howard Fosdick 박사에게 의해 1992년 발표된 "The Sociology of Technology Adaptation" 논문에서 제시한 새로운 기술의 채택모델을 기반으로 현재 워크플로우 마이닝 기술의 국내외적인 기술개발 및 그의 채택현황을 표현한 것이다. 그림 1에서 나타내었듯이 국내외의 모두에서 워크플로우 마이닝에 대한 기술개발은 새로운 기술의 첫 단계인 연구 및 광고 단계에 있어 워크플로우 마이닝 기술의 대중적 인지도 측면만 아니라 그의 적용사례 측면에서도 매우 초보적인 상황이라고 판단된다. 그리고, 선진 외국에서의 워크플로우 마이닝 기술개발은 국내의 기술에 비해 약간의 우위를 보이고 있지만, 차이가 크지 않음을 알 수 있다.

워크플로우 관리 시스템(WFMS: Workflow Management System)을 중심으로 워크플로우 웨어하우징(Workflow Warehousing) 기술, 워크플로우 마이닝(Workflow Mining) 기술에 관한 개념은 데이터 베이스 관리 시스템(DBMS: Database Management System)을 중심으로 한 데이터웨어하우징 기술, 온라인 분석 처리(OLAP: on-line analytical processing)와 데이터 마이닝(Data Mining) 기술로부터 출발했다고도 볼 수 있다. 그리고, 워크플로우 마이닝 기술은 비즈니스 프로세스 리엔지니어링(BPR: Business Process Reengineering), 비즈니스 인텔리전스(BI: Business Intelligence), 비즈니스 프로세스 분석(BPA: Business Process Analysis), 지속적인 프로세스 개선 활동(CPI: Continuous Process Improvement), 그리고 지식 관리(KM: Knowledge Management)와 관련이 있다.

#### 2. 관련연구

##### 2.1 워크플로우 마이닝 기술

현대 비즈니스 프로세스 환경이 복잡화 다양화되어, 실 세계의 프로세스를 반영한 워크플로우의 정의 자체도 복잡화 되었다. 따라서, 워크플로우의 효율적인 개선 모델을 창출해 내기 위하여, 워크플로우 정의 모델과 런타임 데이터를 적절히 활용한 워크플로우 마이닝 기술이 필요하다.

워크플로우 마이닝의 정의를 간단히 살펴보면 다음과 같다.

- 워크플로우 및 마이닝이란 실행 시에 발생한 워크플로우 로그 데이터로부터 정의시점에 알려지지 않은 프로세스의 정보를 추출해 내는 프로세스 발견 방법이다.
- 워크플로우 마이닝은 프로세스를 발견하거나 재발견하는 방법, 그 기술 자체만을 의미하는 것이 아니며, 비즈니스 프로세스의 문제를 이해하고, 이러한 문제를 해결하기 위하여 여러 가지 정보기술을 적용하는 포괄적인 과정을 의미한다.

현재 워크플로우 마이닝(Workflow Mining)은 프로세스 발견 또는 재발견(Process Discovery or Rediscovery)이라고도 불리며, 그 이외에 프로세스 마이닝(Process Mining)이라고도 불린다. 일련의 워크플로우 마이닝 과정은 워크플로우 발견, 워크플로우 디자인, 워크플로우 전개, 워크플로우 실행, 워크플로우 운용, 워크플로우 최적화, 워크플로우 분석 단계의 반복으로 이루어질 수 있으며, 이러한 워크플로우 마이닝의 과정을 거치면서, 워크플로우의 개선, 비즈니스 프로세스 리엔지니어링, 비즈니스 프로세스 변화 관리, 워크플로우 조정 등을 이룰 수 있다.

현재 프로세스 및 워크플로우 마이닝 기술에 대해, 국내외에서는 프로세스를 효율적으로 개선하고자 이미 많은 접근방법을 통해, 연구를 진행 해 나가고 있다. 패턴기반 워크플로우 마이닝, 의존성 기반 워크플로우 마이닝, 이벤트 기반 로그를 통한 워크플로우 마이닝 등 본 고를 통해서 자세한 언급을 해 줄 수 없는 방법들이 많이 있다. 실제 본 연구는 워크플로우 의존성 모델을 중심으로 한 의존성 기반 워크플로우 마이닝 방법과 실행정보를 기초로 한 이벤트 기반 로그를 통한 워크플로우 마이닝 방법을 기초로 진행되었다.

2.2 ICN(Information Control Nets)

워크플로우 마이닝을 위한 제어경로 분석 알고리즘을 구현하기 위해 본 연구에서는 모델의 정의와 분석에 용이한 ICN(Information Control Nets) 모델을 사용한다. ICN은 사무환경에서의 정보의 흐름을 기술하고 분석하기 위해 제안된 모델이다. ICN은 이해가 쉬운 구성 액티비티의 기술을 산출하고 어떤 결합 또는 비일관성을 유발하는 잠재적인 요소의 검출을 위한 평가를 한다. 비즈니스 업무 관련 정보의 흐름을 추적하며 그에 따른 가능한 재구성 구조를 제시한다. ICN 제어흐름 그래프는 큰 원으로 표현되는 일련의 액티비티와 작고 빈 원으로 표현되는 OR노드, 작고 채워진 원으로 표현되는 AND노드, 그리고 이러한 노드들을 연결하는 선(edge)로 구성된다. 화살표(Arc)는 실선(Solid)과 점선(Dashed)으로 표현되는데 이들은 노드들간의 선후관계 및 자료저장소와의 입출력을 표현한다. 비즈니스 프로세스를 정의하는데 있어 다른 비즈니스 프로세스 정의의 언어들을 배제하고, ICN을 활용한 이유로 결정적 역할을 한 이유로는 ICN이 정형화 방법으로 비즈니스 프로세스 모델을 정형적으로 정의를 할 수 있으며, 그로 인해 모델이 타당한지를 판단 할 수 있다는 점 때문이다.

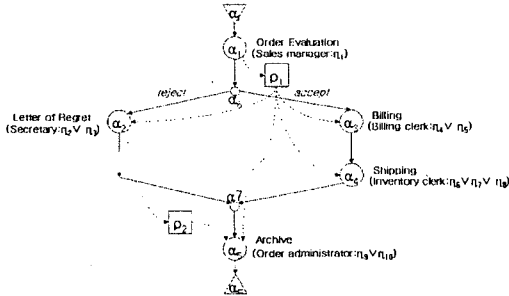


그림2. 주문처리 관계를 나타내는 워크플로우

여기서 전체 프로시저의 이름은 Order Process이며, 각 과정을 담당하는 액티비티는 Order Evaluation, Letter of Regret, Billing, Shipping, Archive( $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ )이다. 그리고  $\alpha_6, \alpha_7$ 은 OR분기를 나타내는 노드이며,  $\rho_1, \rho_2$ 는 저장소를 나타낸다. 또한 직선형태로 된 실선은 제어의 흐름을 나타내고 곡선형태의 실선은 정보의 흐름을 나타낸다.

비즈니스 프로세스를 정의하는데 있어 다른 비즈니스 프로세스 정의 언어들을 배제하고, ICN을 활용한 이유로 결정적 역할을 한 이유로는 ICN이 정형화 방법으로 비즈니스 프로세스 모델을 정형적으로 정의를 할 수 있으며, 그로 인해 모델이 타당한지를 판단 할 수 있다는 점 때문이다.

기본적인 ICN의 구성은  $\Gamma = (\delta, \gamma, \epsilon, \pi, \kappa, I, O)$  과같이 표현되며, 각각의 정의는 다음과 같다.

- $\delta = \delta_1 \cup \delta_0$  이때  $\delta_0 : A \rightarrow \wp(A)$ 은 하나의 액티비티를 후행하는 액티비티들의 집합에 연결하는 관계를 나타내며  $\delta_1 : A \rightarrow \wp(A)$ 은 하나의 액티비티를 선행하는 액티비티들의 집합에 연결하는 관계를 나타내는 관계이다.
- $\gamma = \gamma_1 \cup \gamma_0$  이때  $\gamma_0 : R \rightarrow \wp(A)$ 은 하나의 액티비티를 후속하는 액티비티 집합들을 출력 저장소들의 집합과 연결하는 것 중 하나이며,  $\gamma_1 : R \rightarrow \wp(A)$ 은 하나의 액티비티를 선행하는 액티비티 집합들을 입력 자료저장소들의 집합과 연결하는 관계를 나타내는 것 중 하나이다.
- $\epsilon = \epsilon_a \cup \epsilon_p$  이때  $\epsilon_a : P \rightarrow \wp(A)$ 는 하나의 액티비티를 조합된 역할들의 집합에 액티비티를 단일 값으로 연결하는 것이고,  $\epsilon_p : A \rightarrow \wp(P)$ 는 조합된 액티비티들의 집합을 역할에 단일 값으로 연결하는 것이다.
- $\pi = \pi_p \cup \pi_c$  이때  $\pi_a : C \rightarrow \wp(P)$  조합된 수행자들의 집합을 역할에 단일 값으로 연결하는 것이고,  $\pi_p : P \rightarrow \wp(C)$ 는 조합된 역할들을 수행자에 단일 값으로 연결하는 것이다.
- $\kappa = \kappa_i \cup \kappa_o$  이때  $\kappa_i$ 는  $\alpha \in A$ 일 때,  $(\delta_i(\alpha), \alpha)$  사이에서, 제어흐름(T) 조건들의 집합이고,  $\kappa_o$ 는  $\alpha \in A$ 일 때,  $(\alpha, \delta_o(\alpha))$  사이에서, 제어흐름(T) 조건들의 집합이다. 이 때 집합  $T = \{default, or(conditions), and(conditions)\}$ .
- I 는 초기 입력되는 자료 저장소들의 유한 집합이며, ICN의 실행 전에 어떠한 외부의 프로세스에 의해서 로드 되어야 한다.
- O 는 마지막으로 출력되는 자료저장소들의 유한 집합이며 ICN의 실행 후에 어떠한 외부의 프로세스에 의해서 이용되는 정보를 포함하고 있다고 가정한다.

3. 워크플로우 마이닝을 위한 제어경로분석 알고리즘

워크플로우 프로세스는 일반적으로 복잡한 모델로 정의되지만, 실제로 워크플로우 관리 시스템을 통해 운용될 때에는 하나의 시퀀스를 가진 워크케이스로서 실행된다. 그리고, 각 시퀀스를 갖는 워크케이스는 각 제어경로에 따라 결정되며, 각 워크케이스의 유형을 만들어 낼 수 있다. 하지만, 워크케이스의 유형이 모두 일정한 비율로서 실행되거나 운용되는 것은 아니다. 즉, 모델링 시점에서 정의된 정보와 실행 정보와의 불일치를 찾아 낼 수 있으며, 향후 워크플로우 모델을 개선하고자 할 때, 시스템의 운용정보에 따른 정확한 개선 모델을 만들어 내는데 활용할 수 있다.

워크플로우 프로세스는 워크케이스의 유형별로 분류해 볼 수 있으며, 이는 제어경로의 분석을 통해 이미 정의 시점에 모델 정보로부터 알고리즘을 통해 분석 해 낼 수 있다.

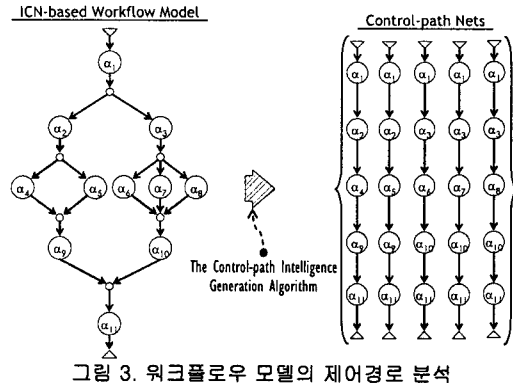


그림 3. 워크플로우 모델의 제어경로 분석

다음 표 1은 제어경로를 통해 실제 실행 시점에서 발생하는 워크케이스의 유형 파악을 위한, 제어경로 분석 알고리즘이다.

표 1. 제어경로 분석 알고리즘

```

Input An Information Control Net (ICN),  $I_1 = (\delta, \gamma, \lambda, \epsilon, \pi, \kappa, I, O)$ ;
Output A Set of Control-path Nets (CpNs),  $\forall W = (\rho, \kappa, I, O)$ ;
Initialize CpN  $\leftarrow \{\varphi\}$ ; /* The empty net of CpN. */
PROCEDURE TRAVERSE(In  $s \leftarrow \{\alpha_1\}$ , CpN) /* Recursive Call. */
BEGIN
   $v \leftarrow s$ ; CpN.Acp  $\leftarrow$  CpN.Acp  $\cup \{v\}$ ;
  WHILE  $((u \leftarrow \delta_o(s)) \neq \{\alpha_F\})$ 
  SWITCH (What type of the activity, u, is?) DO
    Case 'serial-type activity':
       $w \leftarrow u$ ; CpN.Acp  $\leftarrow$  CpN.Acp  $\cup \{w\}$ ;
      CpN. $\rho_o(v) \leftarrow w$ ; CpN. $\rho_o(w) \leftarrow v$ ;
      CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ; CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ;
      break;
    Case 'conjunctive-type (AND-split) activity':
       $w \leftarrow u$ ; CpN.Acp  $\leftarrow$  CpN.Acp  $\cup \{w\}$ ;
      CpN. $\rho_o(v) \leftarrow w$ ; CpN. $\rho_o(w) \leftarrow v$ ;
      CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ; CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ;
      FOR (eachof  $\forall a \in \delta_o(u)$ ) DO
         $x \leftarrow a$ ; CpN.Acp  $\leftarrow$  CpN.Acp  $\cup \{x\}$ ;
        CpN. $\rho_o(w) \leftarrow x$ ; CpN. $\rho_o(x) \leftarrow w$ ;
        CpN. $\beta_o(w) \leftarrow \kappa_o(u)$ ; CpN. $\beta_o(w) \leftarrow \kappa_o(u)$ ;
      END FOR
      FOR (each of  $\forall a \in \delta_o(u)$ ) DO
        Call PROCEDURE TRAVERSE(In  $s \leftarrow a$ , CpN);
      END FOR
      break;
    Case 'disjunctive-type (OR-split) activity':
       $w \leftarrow u$ ; CpN.Acp  $\leftarrow$  CpN.Acp  $\cup \{w\}$ ;
      CpN. $\rho_o(v) \leftarrow w$ ; CpN. $\rho_o(w) \leftarrow v$ ;
      CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ; CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ;
      FOR (eachof  $\forall a \in \delta_o(u)$ ) DO
        Call PROCEDURE TRAVERSE(In  $s \leftarrow a$ , CpN);
      END FOR
      break;
    Default: /* OR-join activity or AND-join activity */
       $w \leftarrow u$ ; CpN.Acp  $\leftarrow$  CpN.Acp  $\cup \{w\}$ ;
      CpN. $\rho_o(v) \leftarrow w$ ; CpN. $\rho_o(w) \leftarrow v$ ;
      CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ; CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ;
      break;
  END SWITCH
   $s \leftarrow u$ ;  $v \leftarrow w$ ;
END WHILE
 $w \leftarrow u$ ; CpN.Acp  $\leftarrow$  CpN.Acp  $\cup \{w\}$ ; /* u is equal to  $\alpha_F$ . */
CpN. $\rho_o(v) \leftarrow w$ ; CpN. $\rho_o(w) \leftarrow v$ ;
CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ; CpN. $\beta_o(v) \leftarrow \kappa_o(s)$ ;
PRINTOUT CpN
END PROCEDURE
    
```

제어경로 분석을 위해 사용되는 제어경로 W는 CpN이라 표현하고, CpN = ( $\rho, \kappa, I, O$ )과 같이 표현되며, 다음과 같이 정의된다

- $\rho = \rho_i \cup \rho_o$  이 때,  $\rho_o : A^{\varphi} \rightarrow \wp(\alpha \in A^{\varphi})$  은 하나의 액티비티를 후행하는 액티비티중 선정된 액티비티에 연결하는 관계를 나타내며,  $\rho_i : A^{\varphi} \rightarrow \wp(\alpha \in A^{\varphi})$  은 하나의 액티비티를 선행하는 하나의 액티비티에 연결하는 관계를 나타내는 것이다.
- $\beta = \beta_i \cup \beta_o$  이 때,  $\beta_i(\alpha)$  는  $\tau \in T^{\varphi}$  일 때,  $(\beta_i(\alpha), \alpha)$  사이에서, 제어흐름 조건의 집합이고,  $\beta_o(\alpha)$  는 제어흐름 조건의 집합이다.  $\tau \in T^{\varphi}$  일 때,  $(\alpha, \beta_o(\alpha))$  에서 발생하는 제어흐름 조건의 집합이다. (이 때,  $\alpha \in A_{cp}$ )
- I 은 ICN 과 일치하며, 입력되는 자료 저장소들의 유한집합이다.
- O 는 ICN 과 일치하며, 출력되는 자료 저장소들의 유한집합이다.

4. 결론 및 향후 연구

본 연구에서는 워크플로우 마이닝을 위하여, 워크케이스의 유형을 분류해 내는 워크플로우 제어경로 분석에 관하여 연구하였고, 워크플로우 제어경로 분석 알고리즘에 관하여 기술하였다.

향후에는 본 연구를 통해 얻어진 알고리즘을 바탕으로 하여, 실제 워크플로우 모델을 분석하고, 워크플로우 로그 정보와의 적절한 비교를 통해, 효율적이고, 효과적인 워크플로우 개선 모델을 제시할 수 있도록 해야 할 것이다.

참고문헌

- [1] Kwanghoon Kim and Clarence A. Ellis: Workflow Reduction for Reachable-path Rediscovery. Proceedings of the ICDM2003 WORKSHOP: Foundations and New Directions in Data Mining, Melbourne, Florida, USA, November, 2003
- [2] Min-Jae Park, Kwang-Hoon Kim, Chang-Min Kim, "An Efficient Workcase Classification Method and Tool in Workflow Mining", 4th Annual International Conference on Computer and Information, pp.80-85, 2005.07
- [3] W. van der Aalst and et al. Workflow Mining: Which processes can be rediscovered? Technical Report, Department of Technology Management, Eindhoven University of Technology, 2002.
- [4] Clarence A. Ellis and Gary J. Nutt, "The Modeling and Analysis of Coordination Systems", University of Colorado/Dept. of Computer Science Technical Report, CU-Cs-639-93, Jan 1993
- [5] 박민재, 김광훈, "워크플로우 마이닝 기술 및 연구개발 동향", 한국인터넷정보학회 학회지, 제6권 제1호, pp.75-81, 2005.03

Acknowledgement

본 연구는 정보통신연구진흥원 정보통신 기초기술연구지원사업 (03-기초-0005)의 지원으로 수행되었음.