

B2C 서비스 산업의 프로세스 마이닝에 대한 연구

강민식*

*남서울대학교 산업경영공학과

A Study on Process Mining for B2C service industry

Min-Shik Kang*

*Dept. of Industrial and Management Engineering , Namseoul University

E-mail : mskang@nsu.ac.kr

요 약

최근 B2C 서비스산업에 있어 기업 간의 경쟁이 심화되고 새로운 비즈니스 가치 창출을 위한 필요성이 증대되고 있는 상황에서, 기업들은 비즈니스 프로세스 관리 기술에 많은 관심을 기울이고 있다. 프로세스의 최적화를 통해 지속적으로 서비스 품질을 개선하기 위해 비즈니스 프로세스 재설계의 근거로 사용될 수 있는 비즈니스 프로세스 마이닝이 중요한 개념으로 인식되고 있다. 하지만 기존의 프로세스 마이닝에 관한 연구에서는 완성되어 있는 프로세스 로그를 기반으로 워크플로우 기반의 프로세스 모델을 추출하는 단조로운 형태였기 때문에 다양한 형태의 비즈니스 프로세스를 표현하는데 한계가 있었다. 본 논문에서는 컨벤션, 대학,병원등 광범위한 지식서비스 분야에서 적합한 Prototype 기관을 Test bed로 다양한 프로세스 마이닝 기법으로 분석하여 해당 조직의 문제 프로세스를 발견하고 개선점을 제안한다. 또한 B2C 서비스 산업에서 적절한 Test bed를 선정하여, 실제 프로세스를 기존의 legacy system의 event log file에서 분석하여 bottle neck process를 찾아내고, 문제 프로세스를 개선하는 과정을 자동화된 모델링 및 분석 툴을 사용하여 실증적으로 보여준다.

키워드

B2C(Business to Customer), Service industry, Process mining

1. 서 론

최근 B2C 서비스산업에 있어 기업 간의 경쟁이 심화되고 새로운 비즈니스 가치 창출을 위한 필요성이 증대되고 있는 상황에서, 기업들은 비즈니스 프로세스 관리 기술에 많은 관심을 기울이고 있다. 기업의 경영혁신의 일환으로 PI와 BPM의 연계를 통한 성과관리를 도입하고 있으며, 고객 서비스 프로세스 개선에 적극적인 '프로세스 혁신'을 시도하고 있다. 또한 실시간 비즈니스 프로세스 및 성과측정, 유연한 프로세스 관리 등 프로세스 관리의 중요성이 커지면서 BPM과 BAM이 시장에서 크게 각광을 받고 있고 매년 30% 성장세를 보이고 있으며 주로 통신, 공공, 금융업계가 주류를 이루고 있다. 그러나 GDP 기준 60%를 상회하는 지식 서비스 분야는 그 광범위한 업종과 규모나 관리 수준의 편차로 인해 프로세스 개선의 사각지대에 있는 실정이다.

이러한 문제점을 해결하기 위하여 비즈니스 프로세스 재설계의 근거로 사용될 수 있는 비즈니스 프로세스 마이닝이 중요한 개념으로 인식되고 있다. 하지만 기존의 프로세스 마이닝에 관한 연구에서는 완성되어 있는 프로세스 로그를 기반으로 워크플로우 기반의 프로세스 모델을 추출하는 단조로운 형태였기 때문에 다양한 형태의 비즈니스 프로세스를 표현하고 적용하는데 한계가 있었다.

본 연구에서는 프로세스 마이닝 기법을 분석하여 가장 효율적인 방법을 이용하여 컨벤션,병원, 대학 등 광범위한 지식서비스 분야에 적용하여 적합한 프로토타입기관을 test bed로 분석하여 해당 조직의 문제 프로세스를 발견하고 개선점을 제한한다. 또한 B2C 서비스 산업에서 적절한 test bed를 선정하여, 실제 프로세스를 기존의 legacy system의 event log file에서 분석하여 bottle neck process를 찾아내고, 문제 프로세스를 개선하는 과정을 자동화된 모델링 및 분석 툴을 사

용하여 그 효율성을 보여준다.

II. 관련 연구

본 논문에서는 표1에 나타나있는 대표적인 프로세스 마이닝 기법인 알파알고리즘, 휴리스틱 알고리즘 그리고 퍼지 알고리즘 3가지 모델 추출 알고리즘 중에서 가장 효율적이라고 알려져 많이 사용하고 있는 휴리스틱 마이닝(Heuristic Mining) 기법을 벤치마킹하여 제안하는 알고리즘과 비교 분석한다. 따라서 휴리스틱 마이닝에 대해 좀 더 자세히 설명한다.

표 1. 프로세스마이닝 기법 장단점 분석

장점	단점
알파 알고리즘	<ul style="list-style-type: none"> • 모든 Log를 동일한 비중으로 분석하여 일반적인 패턴을 도출 • Noise와 완전하지 않은 Log에 약함 • 짧은 Loop를 찾아내기 힘들 • 중첩되거나 내재된 작업을 찾기 힘들 • 복잡한 경로를 가진 프로세스를 찾기 힘들 • 현재 개발된 기법으로는 의미 있는 연결 관계를 찾기 힘들
휴리스틱 알고리즘	<ul style="list-style-type: none"> • 기준값(Threshold)을 설정함으로써 일부 Noise를 제거할 수 있음 • Noise와 완전하지 않은 Log에 덜 민감 • 일반적인 프로세스 모델을 찾는 데 유용 • Heuristic net으로 빈도수를 바로 알 수 있음 • 짧은 Loop를 찾아내기 힘들
퍼지 알고리즘	<ul style="list-style-type: none"> • 복잡한 프로세스를 계층 모델을 통해 보여주어 중요도에 따른 공정을 알기에 용이 • 클러스터를 통한 간략화로 전체 프로세스의 큰 그림을 볼 수 있음 • 의미있는 정보에 집중할 수 있음 • 적절한 수치값들을 결정하기가 어려움

휴리스틱 프로세스 마이닝은 알파 알고리즘보다 불완전한 로그 데이터와 노이즈 (Noise)에 둔감하게 반응하여 프로세스의 신뢰성을 높인다.

프로세스 마이닝은 완벽한 정보를 전제로 하며 완벽한 정보는 시작된 로그는 반드시 완료되어야 하며, 로그에는 노이즈가 존재하지 않는 것을 전제로 한다. 그러나 대부분은 노이즈가 존재하며 완료되지 않은 로그에서는 event 사이의 관계를 정의하기 어렵다. 노이즈를 포함한 로그에서 한 가지의 잘못된 정보는 올바른 결론을 유도하는 데 악영향을 끼칠 수 있다. 휴리스틱 마이닝 기법은 이러한 결함들을 보완한 방법이다. 또한 휴리스틱 프로세스 마이닝은 짧은 고리 구조 (Short loops)를 포함한 알파 알고리즘의 한계점을 보완한 방법으로 기본적으로 네 가지의 순서 관계를 가지고 있다.

- A와 B는 event일 때, 네 가지의 순서 관계는 다음과 같이 정의된다.
- A >B: 워크플로 로그에서 A 발생 후에 직접적으로 B가 발생하면 A>B 관계임. 인과관계 (Causality relation)이라고도 함.
- A →B: A가 선행되어야만 B가 발생할 수 있음. 소위 의존관계 (Dependency relation)라고 하며, 'B가 A에 의존한다.'고 할 수 있음. (A>B 이고

B>A는 아님)

- A#B: A와 B 사이에 직접적인 관계가 없음. 비평행관계(Non-parallel relation)라고 하며, 'A와 B 사이에 의존이 없다.'고 할 수 있음. (A>B 아니고 B>A도 아님)

- A ||B: 분기 (Spilt)와 병합 (Join) 관계를 파악하기 위해 사용됨. A와 B는 평행관계 (Parallel relation)라고 하며 '잠재적인 평행현상을 의미한다.'고 할 수 있음. (A>B 이고 B>A 이다)

또한 세 단계에 걸쳐서 이루어진다. 세 단계는 차례로 'D/F-table (Dependency/Frequency table)의 만들기', 'D/F-table로부터 D/F-graph 만들기', 'D/F-table와 D/F-graph로부터 WF-net을 보완하여 만들기'이다. 휴리스틱 마이닝 기법은 프로세스 마이닝 기법 중에서 가장 정확하고 높은 신뢰도로 프로세스를 유추해낼 수 있으며 이루어진 실험의 모든 상황에서 대부분의 의존관계와 분기 및 병합 관계를 올바르게 발견된다. 일례로, 50%의 노이즈가 있는 로그에서도 올바르게 프로세스를 유추한 결과를 얻을 수 있으며 trace의 수와 각각의 준비된 event가 완료될 확률에 관계없이 탄탄하게 프로세스를 유추할 수 있다.

이 기법에서 일어난 대부분의 오류는 short loops과 관련되어 있으며 short loops과 관련하여 휴리스틱 마이닝의 개선이 필요하며 휴리스틱 룰에서 사용되는 한계치의 크기에 따라 결과가 달라질 수 있으므로, 올바른 결과가 나오도록 적절하게 설정하는 것도 중요하다.

III. 지식서비스 산업 적용 사례

기존의 프로세스 마이닝 기법과 Social network의 handover of work 기반으로 하여 프로세스 분석하여 알고리즘 개발했으며 알고리즘에 대한 핵심 수식은 아래 그림과 같다. 이것은 Handover

정의: Handover of work metrics Let L be a log. For $p_1, p_2 \in P$ and some β ($0 < \beta < 1$):

$$\begin{aligned}
 - p_1 >_L p_2 &= (\sum_{c \in L} |p_1 >_c p_2|) / (\sum_{c \in L} |c| - 1) \\
 - p_1 \dot{>}_L p_2 &= (\sum_{c \in L} \wedge p_1 > p_2} 1) / |L| \\
 - p_1 >_{\beta}^L p_2 &= (\sum_{c \in L} \sum_{1 \leq n < |c|} \beta^{n-1} |p_1 >_n^c p_2|) / (\sum_{c \in L} \sum_{1 \leq n < |c|} \beta^{n-1} (|c| - n)) \\
 - p_1 \dot{>}_{\beta}^L p_2 &= (\sum_{c \in L} \sum_{1 \leq n < |c|} \wedge p_1 > p_2} \beta^{n-1}) / (\sum_{c \in L} \sum_{1 \leq n < |c|} \beta^{n-1}) \\
 - p_1 \geq_L p_2 &= (\sum_{c \in L} |p_1 \geq_c^L p_2|) / (\sum_{c \in L} |c| - 1) \\
 - p_1 \dot{\geq}_L p_2 &= (\sum_{c \in L} \wedge p_1 \geq p_2} 1) / |L| \\
 - p_1 \geq_{\beta}^L p_2 &= (\sum_{c \in L} \sum_{1 \leq n < |c|} \beta^{n-1} |p_1 \geq_n^L p_2|) / (\sum_{c \in L} \sum_{1 \leq n < |c|} \beta^{n-1} (|c| - n)) \\
 - p_1 \dot{\geq}_{\beta}^L p_2 &= (\sum_{c \in L} \sum_{1 \leq n < |c|} \wedge p_1 \geq p_2} \beta^{n-1}) / (\sum_{c \in L} \sum_{1 \leq n < |c|} \beta^{n-1})
 \end{aligned}$$

of work matrix 에 대한 정의이다.

알고리즘의 특징으로는 프로세스 이벤트 로그를 읽을 때 생성이 되는 작업 관계 테이블을 이용하거나 하지 않을 수 있으며, 작업간의 상관관계 계산에 다양한 옵션을 사용하여 보다 유연한 관계 도출이 가능하다. 또한 Threshold 값을 사용하여, 프로세스 모델의 정확도를 조정하는 것이 가능하다.

1. 프로세스 추론 프로토타입 개발

- ProM에 플러그인 형태로 툴 개발

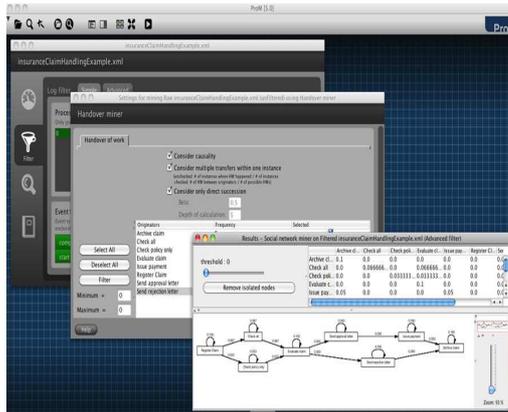


그림 1. 메인 화면

그림1은 메인화면이고 그림2는 마이닝 옵션초기 화면으로, consider causality는 프로세스 마이닝 툴에서 제공하는 causality matrix의 사용 여부를 묻는 것으로, 이를 사용하면 ProM framework에서 제공하는 matrix를 프로세스 모델 도출에 활용한다. “consider multiple transfers within one instance”는 작업 간의 관계 도출에 있어 하나의 케이스에서 여러 번 관계가 관찰될 경우 고려 여부를 표시한다. “consider only direct succession”은 작업간의 관계 도출 시에 거리가 1 이상인 작업의 관계를 고려할지 여부 표시한다.

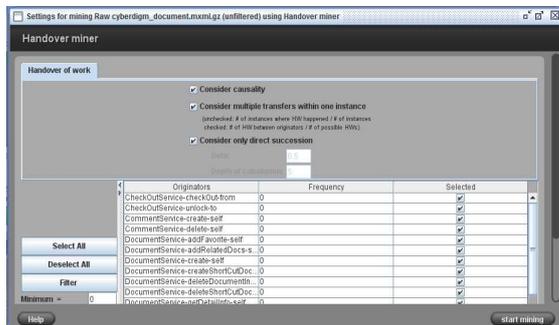


그림 2. 마이닝 옵션 화면

그림3 화면은 마이닝 결과 화면으로 단위 작업의 관계 테이블과 이를 바탕으로 한 프로세스 모

델이 보인다. 관계 threshold 값의 설정을 통해서, 원하는 프로세스 모델의 도출이 가능하다.

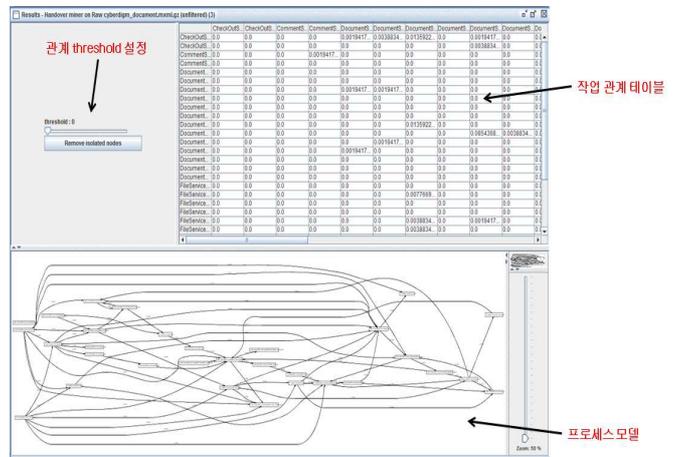


그림 3. 결과 화면

도출된 프로세스의 분석을 위해 프로세스 분석 모듈을 개발하였다. 분석 모듈을 통해서, 도출된 프로세스 상에서 각 작업의 중요도와 주요한 프로세스 흐름의 분석이 가능하다.

2. 적용사례 분석결과

제안한 알고리즘에 의해 다음 사례에 적용하여 분석한 결과이다.

①컨베션 데이터 분석: 프랜차이즈 박람회 에 대한 로그를 분석하였다. Case 376개와 Activity 115개, Event 8493개를 가진 데이터에서 프로세스 모델을 도출하였고 Handover Mining 결과 0.93의 적합도를 얻었고, 기존의 Heuristic Mining으로 0.39의 결과를 얻었다.

②병원 데이터 분석: A대학병원의 데이터 로그를 이용해서 분석하였다. Case 900개 Activity 49개, Event 126,965개를 가진 데이터에서 프로세스 모델을 도출하였고 Handover Mining 결과 1.0의 적합도를 얻었으며 기존의 Heuristic Mining으로 0.74의 결과를 얻었다.

③수강정보 데이터 분석: B대학교의 수강 정보 로그를 이용해서 분석하였다. Case 140개와 Activity 305개, Event 16875개를 가진 데이터에서 프로세스 모델을 도출. Handover Mining 결과 0.99의 적합도를 얻었고, 기존의 Heuristic Mining으로 0.02의 결과를 얻었다.

아래 표와 그래프는 각 분석에 대한 요약된 로그 정보와 적합도 분석 결과이다. 모든 데이터에서 Handover Mining이 기존의 Heuristic Mining보다 높은 적합도 결과 값을 보여준다.

표 2. 적용사례 로그 분석 결과

로그명	로그정보	적합도 결과	
		Heuristic Mining	Handover Mining
컨벤션	Case : 376개 Activity : 115개 시작시간: 2012.3.15 11:37:00 종료시간: 2012.3.17 19:35:00	0.39	0.93
수강정보	Case : 140개 Activity : 305개 시작시간 2008.2.18 11:37:00 종료시간:2012.3.17 19:35:00	0.02	0.99
병원	Case : 900개 Activity : 49개	0.08	1.00

참고 문헌

[1]이승훈외, '로그 기반의 프로세스 마이닝을 통한 BPMN 프로세스 모델 추출', 한국정보과학회 2009 한국컴퓨터종합학술대회 논문집 Vol.36, No.1 pp.137-142, 2009.

[2]김광복외, '프로세스 마이닝에서의 효율적인 적합성 판단 기법', 한국정보과학회 2010 한국컴퓨터종합학술대회 논문집 Vol.37, No.1, pp.66-72, 2010.

[3]W.M.P. van der Aalst and M. Song 'Business process mining: An Industrial application' ELSEVIER Information System 32 pp.713-732, 2007.

[4]C.W.Gunther and W.M.P van der Aalst 'Fuzzy Mining-Adaptive Process Simplification Based on Multi-Perspective Metrics', International Conference on Business Process Management Vol.4714 of Lecture Notes in Computer Science, pp.328-343 Springer-Verlag, Berlin, 2007.

[5] R.S.Mans M.H. Schonenberg and M. Song, 'Process Mining in Health Information', pp.118-125, IEEE Computer Society, 2008.

[6] 송민석, 'c-MES 로그 수집 방안에 대한 연구' 2012 한국 CAD/CAM 학회 학술발표회 논문집 연구, 2012.2.

IV. 결 론

B2C서비스 산업에 있어 최근 기업간의 경쟁이 심화되고 있는 가운데 비즈니스 관리기술에 많은 관심을 기울이고 있으며 ERP(Enterprise Resource Planning), SCM(Supply Chain Management) 등 프로세스 인식 정보시스템들이 널리 쓰이게 되면서 프로세스 마이닝에 대한 연구가 활발하게 이루어지고 있다. 프로세스 마이닝은 프로세스가 실행되는 동안 저장된 이벤트 로그로부터 정보를 추출하는 기법이다. 추출된 로그정보는 비즈니스 프로세스의 분석 및 재설계에 사용될 프로세스 모델을 생성하게 된다. 프로세스 마이닝 기법은 프로세스의 자동화 및 기업의 업무정보들을 관리하는 프로세스 기반 정보시스템의 정확성 및 효율성을 위한 중요한 부분을 차지하지만 현재까지의 연구는 생성된 이벤트 로그로부터 프로세스 모델을 재설계하는 프로세스 발견 기법 (Process Discovery Technique)을 적용한 부분에서만 활발히 진행되었다. 본 논문에서는 기존의 프로세스 마이닝 기법을 기반으로 한 Handover 알고리즘을 이용하여 컨벤션 산업, 대학의 수강정보시스템, 대형병원의 Transaction을 각각 분석한 결과에 대한 요약된 로그 정보와 적합도 분석 결과를 보여주었으며 모든 데이터에서 기존의 Heuristic Mining보다 높은 적합도 결과 값을 보여주고 있다. 이러한 결과는 프로세스 마이닝 알고리즘을 다양한 서비스분야에 적용하여 기존에 컨설팅을 인터뷰에 의해 분석하던 것을 실제 데이터에 의한 정확한 사실(Facts)에 기반한 프로세스 모델을 자동으로 도출해내어 부정확한 인터뷰 기반의 컨설팅의 약점을 보완 하는데 크게 기여할 수 있음을 실증적으로 보여주는 것이라 할 수 있다.