

빅데이터 환경에서 프로세스 마이닝을 이용한 영업감사 상시 모니터링 강화에 대한 연구[☆]

A Study on Continuous Monitoring Reinforcement for Sales Audit Using Process Mining Under Big Data Environment

유 영 석¹ 박 한 규¹ 백 승 훈¹ 홍 성 찬^{*}
Young-Seok Yoo Han-Gyu Park Seung-Hoon Back Sung-Chan Hong

요 약

빅데이터 환경 하에서 프로세스 마이닝은 업무수행 시 발생하는 수많은 데이터들을 활용하여 기업의 ERP시스템 상의 이벤트 로그로부터 프로세스의 수행과 개선에 관련한 많은 정보 및 통찰력을 얻게 해준다. 최근에는 프로세스 마이닝의 최대 강점을 활용하여, 기업조직의 감사업무에 적극적으로 활용하려고 하는 다양한 연구 활동이 활발히 진행 중에 있다. 그러나 프로세스 마이닝을 이용한 영업감사 적용에 관한 최근의 국내 연구는 빅데이터 환경 하에서는 매우 미흡한 실정이다. 이러한 상황에 착안해서, 본 연구는 프로세스 마이닝이 감사 분야에 적용된 기존 연구를 한층 더 강화시킴으로써, 온라인 방식 및 전통적 감사에 최적으로 적용할 수 있는 프로세스 마이닝 적용 방안을 제안하고자 한다. 또한 빅데이터 환경하에서 기업 조직의 리스크 발생 요인들을 사전에 모니터링함으로써, 리스크의 조기 발견 및 이를 예방할 수 있는 상시 모니터링 정보 서비스 시스템을 제안한다. 이를 위해서 리스크 요인을 기반으로 데이터들을 추출하고 평가에 대한 관리기준을 설정한다. 본 논문의 연구의 범위는 영업감사에 있어 실제 사례를 통해 위험요소의 사전 검증 시스템을 설계 한다. 그리고 이러한 영업감사 시스템을 통해 예방감사 실현, 높은 리스크 요인들에 대한 상시 대응, 사기 발생 억제, 규칙 및 지침 위반에 대한 적시조치, 경영환경 변화에 대한 감사 아이템 추가 발굴 체계 구축, 프로세스 개선 중심의 사전적인 컨설팅 감사의 실현, 내부통제 회계제도 준수 및 강화를 행한다. 이 결과로 영업 감사 실시간 관련 통합 모니터링이 강화되어 재무 리스크 회피, 감사기간 단축 및 감사 품질 개선 등의 효과가 나타났다.

☞ 주제어 : 프로세스 마이닝, 내부 감사, 실시간 모니터링, 빅데이터, 상시 모니터링

ABSTRACT

Process mining in big data environment utilize a number of data were generated from the business process. It generates lots of knowledge and insights regarding implementation and improvement of the process through the event log of the company's enterprise resource planning (ERP) system. In recent years, various research activities engaged with the audit work of company organizations are trying actively by using the maximum strength of the mining process. However, domestic studies on applicable sales auditing system for the process mining are insufficient under big data environment.

Therefore, we propose process-mining methods that can be optimally applied to online and traditional auditing system. In advance, we propose continuous monitoring information system that can early detect and prevent the risk under the big data environment by monitoring risk factors in the organizations of enterprise. The scope of the research of this paper is to design a pre-verification system for risk factor via practical examples in sales auditing. Furthermore, realizations of preventive audit, continuous monitoring for high risk, reduction of fraud, and timely action for violation of rules are enhanced by proposed sales auditing system. According to the simulation results, avoidance of financial risks, reduction of audit period, and improvement of audit quality are represented.

☞ keyword : Process Mining, Internal Audit, Real Time Monitoring, Big Data, Continuous Monitoring

1. 서 론

¹ Department of Information and Telecommunications, Hanshin University, Osanshi 447-791, Korea

^{*} Corresponding author (schong@hs.ac.kr)

[Received 6 June 2016, Reviewed 21 June 2016(R2 25 July 2016), Accepted 9 October 2016]

☆ 본 논문은 한신대학교 학술연구비 지원에 의하여 연구된 논문임

IT화 및 글로벌화가 급속히 진행되고 있고 국가간 정보의 장벽이 무너져 디지털 경제가 급속히 확산됨으로써 그 규모를 측정할 수 없을 정도로 대규모 정보가 생산되는 '빅데이터(Big Data)' 환경이 도래하였다. 미래 경쟁력을 좌우하는 핵심 자원인 빅데이터는 기업 및 공공 부문도 이에 대한 정확한 이해와 이를 효과적으로 활용하기 위한 전략 정책이 매우 시급하다[1,2].

빅데이터란 일상적인 디지털 환경하에서 생성되는 정형 및 비정형 데이터로 그 규모가 매우 방대하고, Life Cycle도 짧고, 형태도 수치 데이터뿐만 아니라 문자, 이미지, 영상 데이터를 포함하는 대규모 데이터를 말한다. 빅데이터 환경은 지난 과거에 비해 다양한 수집 방법과 수집되는 데이터의 양이 폭증했다는 점과 함께 데이터의 종류도 매우 다양해짐에 따라 사람들의 행동패턴은 물론 위치정보 및 SNS를 통해 다양한 생각과 의견까지도 분석하고 예측할 수 있다[1,7].

본 논문에서는 프로세스 마이닝이 감사분야에 적용된 기존 연구를 한층 더 강화 시킴으로써, 기업의 영업 리스크 발생요인들을 상시 모니터링하고 예방 관리하기 위하여, 빅 데이터 환경 하에서 온라인 방식 및 전통적 감사에 최적으로 적용할 수 있는 프로세스 마이닝 적용방안을 제안하고자 한다.

2. 프로세스 마이닝의 개요 및 적용분야

2.1 프로세스 마이닝 개요

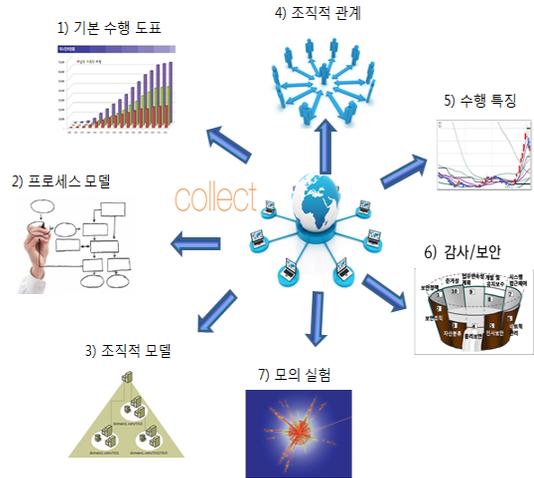
빅데이터 시대를 맞아 프로세스 분석 기법은 새로운 변혁의 시기에 있다. 프로세스 분석 기법은 디지털 경제의 발전으로 인해 기업의 비즈니스 프로세스 및 조직의 업무 프로세스와 밀접하게 관련되어 있는 각종 정보인 BPM, ERP, CRM, SCM 등을 통해서 다양한 기업의 업무 처리 시스템 상의 모든 기록은 이벤트 로그 형태로 기록 관리되고 있다. 프로세스 마이닝은 이와 같이 시스템 상에 기록 관리되어 있는 이벤트 로그를 분석하여 의미 있고 가치 있는 정보를 추출해 내는 것을 목적으로 하는 방법으로 기업 경영 혁신에 있어 필수적인 요소이다.

또한 프로세스 마이닝에서 프로세스에 대한 분석, 프로세스 간소화, 교차점 식별 및 문제 예측, 지정된 업무 수행률 및 지침 위반 검사 등에 대한 프로세스 로그 분석은 다양한 목적으로 연구 대상에 활용될 수 있다[8].

2.2 프로세스 마이닝의 적용 분야

프로세스 마이닝의 목적은 ERP, BPM, SCM, CRM 등 다양한 기업의 업무 처리 시스템에서 기록되는 빅데이터의 이벤트에 대한 로그를 분석하여 유의미한 정보를 찾아내는 것을 의미한다. 아래 그림 1은 프로세스 마이닝의 세부적인 연구 분야들을 나타낸 그림이다. 프로세스 마이닝의 주된 연구 분야는 프로세스 성과 측정, 조직 모델

및 프로세스 모델 도출, 시뮬레이션 분석 등이다[3].



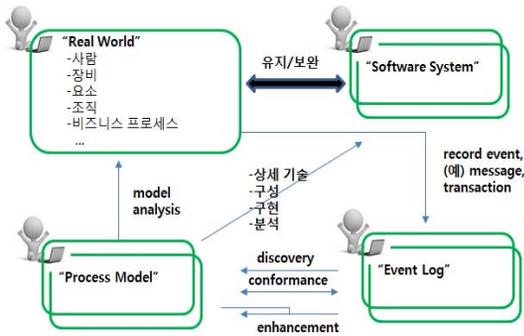
(그림 1) 프로세스 마이닝 연구 분야
(Figure 1) Process Mining Research Areas

2.3 State of the art

프로세스 마이닝의 출발점은 이벤트 로그이며, 이벤트가 발생 순서대로 기록되어 있다고 가정한다. 또한 하나의 이벤트는 프로세스의 단위 작업을 나타내며, 변동 불가능 프로세스와 관련을 맺고 있다. 이벤트 로그는 이벤트에 대한 여러 가지의 정보를 저장할 수 있으며, 프로세스 마이닝 기법에서 이러한 부가적인 정보들을 활용할 수 있다. 예를 들어, 작업 개시나 업무수행자, 시스템, 장비 등과 같은 수행 리소스, 이벤트 발생 시간을 의미하는 이벤트인 타임스탬프, 주문의 규모와 같은 이벤트에 연관된 데이터 등을 활용할 수 있다.

그림 2와 같이, 이벤트 로그는 크게 세 가지 유형의 프로세스 마이닝 실행에 적용할 수 있다. 프로세스 마이닝의 실행 단계 중 첫 번째 유형은 도출(discovery)이며, 이 기법은 사전 정보 없이 이벤트 로그에서 모델을 생성 및 추출하는 기법이다. 매우 중요한 프로세스 마이닝 기법으로는 프로세스 모델 도출이다. 현재 다양한 기관들은 최근에 개발한 프로세스 마이닝 기법을 이용하여 이벤트 로그로부터 적용되고 있는 프로세스를 도출해 내고 있다.

프로세스 마이닝의 두 번째 유형은 적합성(conformance)이며, 그 검사는 보통 기존의 프로세스 모델과, 기존 모델에서 생성된 이벤트 로그를 비교 분석한다[4,13].



(그림 2) 프로세스 마이닝의 세 가지 주요 기법:

㉠프로세스 도출 ㉡적합도 검사 ㉢모델 향상

(Figure 2) Process Mining's Three Main Techniques: ㉠process discovery ㉡conformance checking ㉢model extension

적합성 검사는 로그에 기록된 현실이 모델에 일치하는지와 모델이 현실에 일치하는 것인지 확인하는 데 이용할 수 있다. 적합성 검사는 다음과 같은 모델에 적용할 수 있다. 1) 프로세스 및 조직 모델과 선언적 프로세스 모델 2) 적합성 관련 법규 3) 비즈니스 규칙 및 정책 등 다양한 모델에 적용할 수 있다.

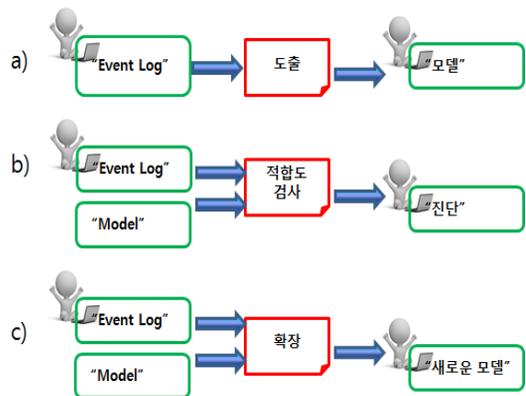
프로세스 마이닝의 세 번째 유형은 향상(enhancement)이며, 그에 대한 기본적인 아이디어는 이벤트 로그상의 실제 프로세스와 관련된 해당 정보를 통해 기존의 프로세스 모델을 변경하고 확장하여 전체적인 내용을 개선하는 것이다. 적합성 검사는 모델과 현실 사이의 정합성을 측정하는 것과는 반대로 향상은 이전 모델을 개선하거나 확장시키는 것을 목표로 한다. 예를 들어, 이벤트 로그의 타임스탬프를 통해 기존 모델을 교착점, 서비스 수준, 업무 처리 시간, 빈도 등을 보여주는 모델로 변경 및 확장할 수 있다[5,13].

그림 3은 입력과 출력 관점에서 프로세스 마이닝의 세 가지 유형을 보여준다. 하나의 이벤트 로그를 입력 받아 그 결과 하나의 새로운 모델을 생성할 수 있는 도출에 대한 기법들이 있다. 이 기법들은 Petri net, BPMN, EPC, UML activity diagram 등이며 프로세스가 도출된 모델이지만, 소셜네트워크 같은 비정형 데이터에 대한 모델이 될 수도 있다. 적합도 검사 기법은 Event Log와 모델을 필요로 하며, 이 결과는 모델과 로그의 공통점과 차이점을 보여주는 정보이다. 이전의 모델을 수정 또는 확장을 의미하는 모델 향상 기법도 존재하고 있는 Event Log와 모델을 필요로 한다[5,13].

결과는 개선된 모델 또는 확장된 모델이며, 프로세스 마이닝은 여러 관점을 다루고 있다. 다음과 같은 관점이 있다.

1) 프로세스 흐름 관점은 작업의 순서와 같은 프로세스 흐름에 초점을 둔다[13].

프로세스 흐름 관점의 목표는 모든 가능한 경로의 특징을 파악하여, 그 결과는 보통 Petri net 모델이나, BPMN, EPC, UML activity diagram과 같은 프로세스 모델로 표현된다.



(그림 3) 입력과 출력 관점에서 본 프로세스 마이닝:

㉠프로세스 도출 ㉡적합도 검사 ㉢모델 향상

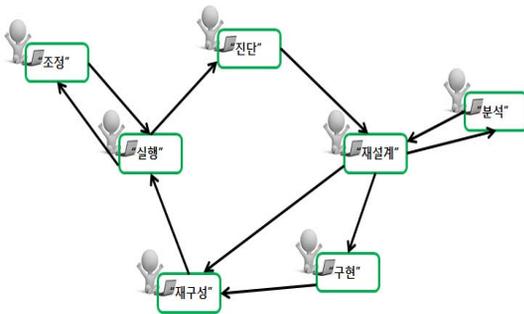
(Figure 3) Process Mining Viewing Input and Output: ㉠process discovery ㉡conformance checking ㉢model extension

2) 이벤트 로그에 감춰진 리소스에 대한 정보에 초점을 맞추어, 시스템, 작업자, 부서, 역할과 같은 어떤 행위자가 업무 수행에 관여하고, 그들이 어떻게 상호 연결되어 있는지를 분석하는 것이 조직 관점이며, 조직 관점 분석의 목적은 역할이나 기능에 따라 구성원들을 분류하여 조직 체계를 만들거나 구성원들 사이의 소셜 네트워크를 보여주는 것이다.

3) 케이스의 특징에 초점을 맞추며, 하나의 케이스는 프로세스 내에서의 경로 또는 케이스에 참여하는 작업자를 통해 특성을 분석하는 것이 케이스적인 관점이다. 보충 주문을 실행하는 케이스가 있으면, 고객으로부터 주문 받은 상품의 수량과 상품공급자는 유용한 정보가 되며, 케이스와 연관된 데이터의 값에 따라서 케이스의 특성을 분석할 수도 있다.

4) 이벤트가 타임스탬프를 포함하면, 교차점 도출, 서비스의 수준 측정, 자원 활용도 모니터링, 수행 중인 케이스의 잔여 시간 예측 등이 가능하며, 이처럼 이벤트의 시간과 빈도를 고려하는 것이 시간 관점이다.

한편으로 프로세스 마이닝에 대한 대표적인 오해가 몇 가지 있다. 일부 기업, 분석가, 연구자들은 프로세스 마이닝의 범위를 단지 오프라인 분석에만 사용되는 프로세스 도출을 위한 특별한 데이터 마이닝 기법으로 생각하나 프로세스 마이닝에서 그림 4는 BPM(Business Process Management) Life_cycle을 통해 프로세스 마이닝의 위치를 보여준다. BPM Life_cycle은 비즈니스 프로세스와 관련 정보시스템의 7단계를 보여 준다[6,13].



(그림 4) Business Process와 관련 정보 시스템의 BPM Life_cycle 7단계

(Figure 4) Business Process and Related Information System's BPM Life_cycle 7 Steps

설계 또는 재설계 단계에서는 기존의 프로세스 모델이 수정되거나 새로운 프로세스 모델이 생성되고, 이 후 분석 단계에서는 후보 모델, 대체 모델들에 대한 분석이 이루어진다. 설계 또는 재설계 단계를 거치고 나면 모델이 구현되거나, 혹은 기존의 시스템이 구성되거나 재구성된다. 실행 단계에서는 설계된 모델이 수행이 되며, 프로세스의 모니터링이 이루어진다. 또한 사소한 수정 사항들은 재설계 단계를 거치지 않고 수정되며 이를 조정 단계라 한다.

진단 단계에서는 실행된 프로세스가 분석되고, 분석 결과를 통해 얻은 정보를 바탕으로 새로운 프로세스를 설계 또는 재설계 한다.

진단 단계는 프로세스 마이닝 기법의 사용이 매우 유용하게 활용되며, 프로세스 마이닝은 그림 4의 대부분의 모든 단계에서 유용하게 사용된다. 한편 활용에 대한 진단 단계에만 국한되지 않으며, 실행 단계에서 프로세스

마이닝의 기법들에 대한 운영적 지원을 위해 이를 활용할 수 있다. 이전 데이터를 기반으로 학습된 모델을 통해 수행 중인 케이스에 영향을 줄 수 있다. 새로운 프로세스의 구성 또는 재구성에 이를 적용하고자 하면 이전의 만들어진 의사 결정이 새로운 프로세스 생성에도 쓰여져야만 한다. 표 1에서는 BPM Life_cycle의 단계별 수행 내용을 보여주고 있다.

그림 5는 프로세스 마이닝 활동과 결과물에 초점을 맞추어 프로세스 마이닝 프로젝트에 나올 수 있는 단계들에 대해서 보여주고 있다[8].

정보 시스템의 0단계에서는 프로세스 마이닝 프로젝트 계획 작성과 계획에 대한 타당성 평가로 시작한다. 그리고 1단계에서는 프로젝트를 시작한 후에 도메인 전문가, 경영진, 시스템으로부터 이벤트 데이터와 목적, 의문 사항, 모델 등을 도출한다. 또한 2단계에서는 분석을 하려면 어떤 자료가 사용될 수 있을까라는 가용 데이터, 어떤 것이 중요한 질문인가라는 도메인으로부터 그림 5의 1단계로부터 추출된 과거 데이터, 핸드메이드 모델, 목적, 의문점 등과 같은 중간 산출물에 대한 이해가 요구된다. 결과적으로 2단계에는 자동화된 프로세스 도출 기법을 적용하여 프로세스 흐름 모델을 도출하고, 이벤트 로그에 연결한다. 기존 프로세스를 설계 또는 재설계 하기 위해서는 도출된 프로세스 흐름 모델의 의문 사항에 대한 결과를 활용하여 이를 바탕으로 해당 프로세스를 재설계하거나 수정할 수 있다.

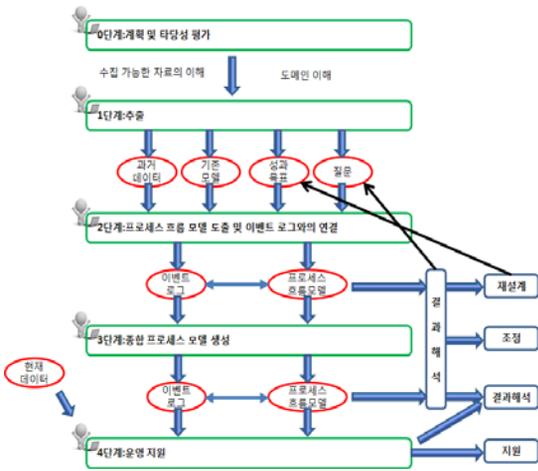
또한 모델을 활용하여 빈도수가 낮은 작업이거나 이상치에 해당하는 케이스를 제거하거나, 누락된 이벤트를 추가할 수 있으며, Event Log를 필터링하거나 수정할 수 있다. 때로는 동일한 프로세스 인스턴스에 속하는 이벤트를 서로 연관시키도록 노력을 해야 한다.

로그에 있는 이벤트는 프로세스 모델의 개체에 연결되어야 하며, 프로세스가 비교적 구조적일 경우 3단계에서는 프로세스 모델을 데이터, 시간, 자원의 관점으로 확장할 수 있다.

각 작업의 대기 시간은 이벤트와 연관된 타임스탬프를 통해 예측될 수 있으며, 2단계에서 생성된 이벤트 로그와 해당 모델의 관계를 통해 새로운 모델을 확장할 수 있다. 이렇게 해당 모델을 확장하여 부가적인 질문에 대한 해답을 찾을 수 있으며, 추가적인 분석이 가능하다.

4단계에서는 운영적 지원의 목적으로 3단계에서 형성된 모델들을 사용할 수 있다. 수행 중인 케이스의 정보와 결합되어진 과거 이벤트 데이터에서 추출된 지식은 수행 중인 프로세스에 대한 개입, 예측, 추천에 사용될 수 있

다. 따라서 3단계와 4단계는 프로세스가 안정적이고 체계적일 때 모델의 성공적인 생성을 이룰 수 있다. 프로세스 마이닝 프로젝트 수행 5단계 life_cycle 모델의 모든 단계를 지원할 수 있는 기법과 도구가 존재하기는 하지만, 현재 프로세스 마이닝은 상대적으로 새로운 유형에 대한 흐름이기 때문에 이를 이용할 수 있는 툴이 부족한 것은 사실이다. 또한 목적 지향적 사용자들은 종종 프로세스 마이닝의 잠재성과 한계성을 인식하지 못하고 있다 [8,13].



(그림 5) 프로세스 마이닝 프로젝트 수행 5단계 라이프 사이클 모델
(Figure 5) Process Mining Project Executing 5 Steps Life cycle model

(표 1) BPM 라이프 사이클 단계별 수행 내용
(Table 1) BPM Life_cycle Each Steps Executing Context

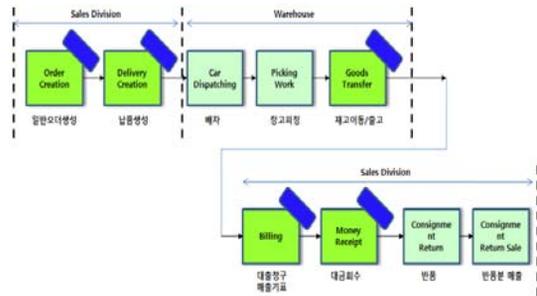
단계 이름	단계별 수행 내용
(재)디자인	-새로운 프로세스 모델 생성 -기존 프로세스 모델 수정
분석	-후보 모델과 대체 모델 비교
(재)구성	-(재)디자인 단계 이후 -기존 시스템에 대한 (재)구성
수행	-디자인된 모델 수행 -프로세스 모니터링
적용	-사소한 수정사항들을 재 디자인 단계를 거치지 않고 수정하며 적용
진단	-생성된 프로세스 분석 -분석 결과를 통해 얻은 정보를 바탕으로 새로운 프로세스 재설계

3. 제안한 프로세스 마이닝

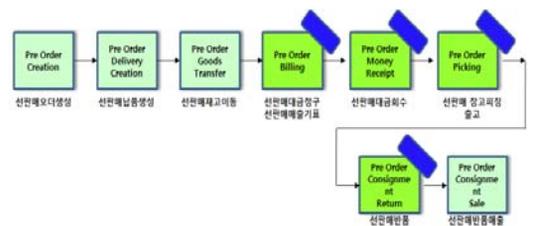
프로세스 마이닝의 목적은 기업의 Business Process에서 일어나는 업무처리를 통해 유익한 정보를 발견하는 것이며, 그 결과로 생성된 프로세스 마이닝은 기업의 Business Process 혁신에 활용될 수 있다[10,11].

그림 6과 그림 7은 이벤트 형식으로의 프로세스 흐름을 보여주고 있다. Business Process상의 모든 업무는 기업 내부의 정보시스템인 CRM, ERP, Workflow, SCM 등에서 수행된다[11,12].

그림 6과 그림 7의 흐름을 바탕으로 의미 있는 정보와 지식을 추출해내는 프로세스 마이닝은 이미 축적되어 있는 Transaction Log Data의 분석을 통해 프로세스를 발견하며 해당 업무 프로세스를 개선한다.



(그림 6) General Sales Order Flow
(Figure 6) General Sales Order Flow



(그림 7) Pre Sales Order Flow
(Figure 7) Pre Sales Order Flow

프로세스 마이닝의 연구는 다음과 같다.

프로세스 마이닝의 연구는 1) 첫째로 업무가 어떻게 진행되는지에 대한 프로세스 관점 2) 둘째로 업무 수행자가 누구인지에 대한 조직 관점 3) 마지막으로 임의 사건에 대한 수행자 혹은 프로세스 흐름에 대한 특성이 무엇인지에 대한 사건 관점 등으로 구분할 수 있다.

세 가지 관점 중 프로세스 관점의 연구가 가장 활발히 진행되고 있다.

Agrawal et al.은 Work_flow 관리시스템의 로그를 기초 자료로 하여 Business Process를 모델링하는 방법을 제시하였으며, Cook and Wolf는 순수 알고리즘, 뉴럴 네트워크, 마코브(Markovian) 방법으로 Process Data를 분석하여 소프트웨어공학에서 프로세스 모델을 개선하는 연구를 하였고 Herbst는 Work_flow 시스템에 기계학습 알고리즘(Machine learning algorithm)을 이용하여 Work_flow 프로세스 모델을 개선한 연구를 하였다. 한편 Aalst et al.은 α -algorithm을 이용하여 선택실행 프로세스를 고려한 Work_flow 프로세스 모델을 재발견(rediscovery)하는 연구를 수행하였고, Aalst et al.와 Medeiros et al.은 선택실행 프로세스뿐만 아니라 동시 실행 프로세스를 고려한 Business Process 개선과 발견을 위한 유전자 알고리즘이 실행되었다.

이들은 유전자 알고리즘을 적용하기 위해 Business Process를 Petri Net으로 표현하여 Casual Matrix를 구성하였고, Petri Net의 점화규칙을 이용하여 프로세스 적응도를 평가하였다. 하지만 이미 체계화된 프로세스 모델을 기반으로 Casual Matrix 초기해를 생성하기 때문에 구조화된 프로세스 모델이 존재하지 않는 경우에는 프로세스 추출이 어렵다. 또한 Petri Net의 점화규칙을 이용하여 해당 프로세스 처리에 대한 도출된 해들의 적응도를 평가하는 시간과 비용(cost)이 상당히 높았으며, 데이터량이 증가할수록 알고리즘의 효과성은 반대로 급격하게 낮아졌다. 그리고 유전자 알고리즘은 정확한 해의 탐색이 어렵기 때문에 Business Process상의 모든 업무와 업무간의 상세한 프로세스 흐름관계를 탐색하고자 하는 데는 어려움이 있다[4,6].

본 연구에서는 A기업 영업부문의 Transaction Log를 대상으로 선택실행 프로세스, 동시실행 프로세스 그리고 중요한 프로세스이지만 수행 빈도가 낮아 이용율이 저조한 Business Process를 검증함으로써 기업의 전반적인 업무상 이상징후들을 보다 정확하게 파악하기 위하여 그림 8과 같이 데이터 마이닝의 전체적인 흐름을 이용한 프로세스 마이닝 알고리즘을 개발하였다.

General_Sales_Order()

/*공장에서 생산한 제품을 영업부서에서 거래처 영업/마케팅 활동을 통하여 수주할 경우 영업오더에 대한 출고지시, 물류배송 및 최종 판매대금 회수 등 일반적인 영업 오더를 대상으로 함*/

{

Order_Creation();

/*영업담당자가 거래처로부터 일반적으로 수주 받은 내용을 시스템상에 입력 처리함 (거래처명, 수주물량, 납기, 단가, 인도처 등)*/

Delivery_Creation();

/*제품생산이 완료되어 공장에 재고가 있는 경우,영업담당자가 발행하는 출하지시서에 의거하여 납품생성을 하는 절차*/

Car_Dispatching();

/*물류센터에서 영업담당자의 출하지시에 따라 운송차량을 배정하고, 피킹지시서 및 운송계획서를 출력함*/

Picking_Work();

/*물류센터에서 피킹지시서에 의거하여 운송제품에 바코드를 부여하여 피킹작업을 실시함*/

Goods_Transfer();

/*물류센터에서 창고 제품출하를 하며 시스템상에 재고이동 처리함*/

Billing();

/*영업담당자는 물류센터에서 제품출하 정보를 확인한 후, 거래처에 대금청구 및 세금계산서 발행하고 매출회계처리를 함*/

Money_Receipt();

/*거래처와 계약된 대금 결제조건에 따라 판매대금을 회수함*/

Consignment_Return();

/*거래처의 사정 혹은 제품 하자로 인하여 거래처로부터 반품받음*/

Consignment_Return_Sale();

/*거래처로부터 반품받은 제품을 가격조정 혹은 시기조정하여 다시 판매함*/

}

Pre_Sales_Order()

/*공장에서 생산한 제품을 영업부서에서 거래처 영업/마케팅 활동을 통하여 수주할 경우, 거래처의 요청에 의거하여 제품을 거래처에 선 판매하고 제품은 거래처의 희망납품시기에 맞추어 배송하는 오더를 대상으로 함*/

{

Pre_Order_Creation();

/*영업담당자가 거래처로부터 수주 받은 선판매오더 내용을 시스템상에 입력 처리함 (거래처명, 수주물량, 납기, 단가, 인도처 등)*/

Pre_Order_Delivery_Creation();

/*회사에 재고가 보유된 상태에서 거래처의 요청에 의거하여 영업담당자가 선판매 오더에 대한 납품생성을 하는 절차*/

Pre_Order_Goods_Transfer();

/*선 판매오더를 시스템상 정상재고에서 선판매재고로 전환함*/

Pre_Order_Billing()

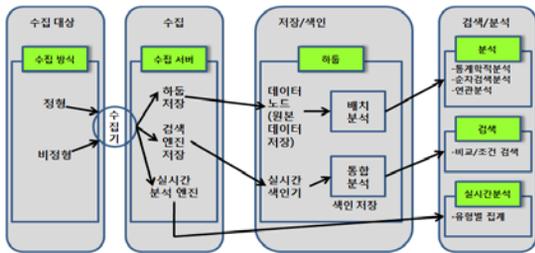
/*선 판매분에 대한 매출 회계처리를 하고 거래처에 대금 청구함*/

```

Pre_Order_Money_Receipt( );
/*거래처와 계약된 선판매대금 결제조건에 따라 판매대금을 회수함*/
Pre_Order_Picking_Work( );
/*거래처의 요청에 의하여 물류부서에서 제품 피킹하여 거래처가 원하는 납품처로 출고함*/
Pre_Order_Consignment_Return( );
/*선 판매제품이 물품 하자 등의 사유로 인하여 거래처로부터 반품받음*/
Pre_Order_Consignment_Sale( );
/*반품받은 선판매제품을 일부 가격조정 혹은 시기조정하여 재판매함*/
}
    
```

(그림 8) "A"사의 프로세스 마이닝 알고리즘
(Figure 8) "A" Company's Process Mining Algorithm

대용량 데이터의 전처리 및 모델링을 수행하는 경우 Index에 의한 병렬 데이터 접근과, In-Memory에 의한 분산 데이터 처리 및 병렬(Parallel) 연산 처리방식을 적용하여 빅데이터 분석을 위한 최적화 기능을 제공한다[9]. 기존의 문제점을 극복하고 다양한 데이터 수집, 실시간 고속 처리, 센서 등 IoT 데이터 처리를 위한 신 개념의 빅데이터 솔루션이 필요하다. 이를 해결하기 위한 본 논문의 전체적인 시스템 개념도는 그림 9와 같다[9].



(그림 9) 시스템 개념도
(Figure 9) System Flow

본 연구에서는 이러한 빅데이터 솔루션을 활용하여 General Sales Order인 경우 표 2의 Sales Order Check Point에 대하여 표 3의 Slaes Order Check Point Control에서와 같이 도착희망일 이후 출고, Order Creation Date후 2일 초과하여 출고, Billing후 익월말 초과하여 대금회수, 가격할인 수정사항 발생 등의 비정상 조건(리스크 요인)들이 도출되는 것을 알 수 있었다.

본 논문의 General Slaes Order와 Pre Sales Order에 대한 Sales Order Check Point는 표 2와 같다.

(표 2) Sales Order Check Point
(Table 2) Sales Order Check Point

	Check Point
General Sales Order	①거래처의 도착희망일과 Goods Transfer ②Order Creation Date와Goods Transfer Date의 차이기간(재고판매인 경우) ③Billing Date와 Money Receipt Date ④Billing시 가격할인 적용
Pre Sales Order	⑤Billing Date와 Money Receipt Date ⑥Pre Sales Order Stock =(Pre Order Billing-Pre Order Picking Work)

본 논문의 General Slaes Order와 Pre Sales Order의 정상 및 비정상조건(Risk 요인)에 대한 Slaes Order Check Point Control은 표 3과 같다.

(표 3) Sales Order Check Point Control
(Table 3) Sales Order Check Point Control

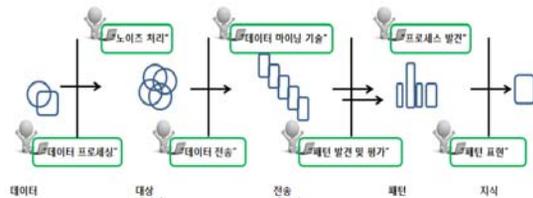
	정상조건	비정상조건 (Risk 요인)
General Sales Order	①도착희망일 이내 출고 ②Order Creation Date후 2일 이내 출고 ③Billing후 익월말내 대금 회수 ④가격할인 수정사항 없음	①도착희망일 이후 출고 ②Order Creation Date후 2일 초과하여 출고 ③Billing후 익월말 초과하여 대금회수 ④가격할인 수정사항 있음
Pre Sales Order	⑤Billing 당월에 입금 ⑥Billing후 익익월말 이내 출고	⑤Billing 당월이후 입금 ⑥Billing후 익익월말 이후 출고

4. 프로세스 마이닝과 빅데이터 분석

프로세스 발견을 위한 프로세스 마이닝 프로세스는 시작 업무부터 마지막 업무까지 수행되는 순서와 절차가 존재하고 다양한 Sub-process들이 포함되어 있으며 Transaction Log에는 업무 트랜잭션이 발생할 때 마다 누적 데이터들이 끊임없이 진행된다.

이와 같이 선택적 처리에서 얻어지는 기업의 로그 데이터에서 비즈니스의 효율적인 수행 프로세스 혹은 이상 징후가 있는 업무 프로세스를 발견하기 위해서는 업무와 업무간의 순서와 연관성을 찾아야 한다. 연관성을 찾기 위해서는 대량의 데이터에서 숨겨진 순서, 이상징후 처리, 선택적 프로세스 선정을 발견하여 의사결정에 활용

해야 한다. 이러한 목적을 달성하기 위하여 데이터 마이닝 기법 중의 하나인 순차 패턴과 연관규칙을 활용하는 기법은 업무 수행상 이상 징후를 찾을 수 있는 효과적인 방법이다.



(그림 10) 데이터 마이닝을 이용한 이상 프로세스 발견
(Figure 10) Error Process Discovery Using Data Mining

연관규칙은 하나의 Transaction내에서 동시에 발생하는 항목간의 연관성을 발견하는 것으로 Business Process 내 동시에 발생하는 업무를 발견하는 데 유용한 기법이며, 순차패턴은 연관규칙에 시간 변이를 추가한 것으로 트랜잭션 내 항목들의 시간적 관계를 찾아 Business Process내의 업무 순서를 발견하는데 유용하다. 이러한 두 기법의 특징을 동시에 고려할 경우 그림 10과 같이 효율적으로 비즈니스 이상 프로세스를 발견할 수 있다. 또한 본 연구는 이와 같은 두 가지 기법을 사용하여 표 3과 같이 모니터링을 강화할 수 있었다.

5. 결 론

본 연구는 전사 차원의 빅데이터 수집 및 분석 환경을 확보하고 시스템 로그, WEB, SNS 등의 통합 분석환경을 구축하였다. 즉, 전사적으로 수집한 각종 데이터 등으로부터 숨겨진 패턴과 연관성 탐색 및 이상 징후 발견을 목적으로 하였다. 그리고 미래 예측을 통한 시스템의 필요 뿐만 아니라 시스템 요구 사항과 이상 징후 발견을 능동적으로 해결하고자 했다.

이에 대한 해결방안으로 본 연구에서는 선택적 처리에서 얻어지는 기업의 로그 데이터에서 비즈니스의 비효율적인 업무 수행 프로세스 혹은 이상징후가 있는 업무 프로세스를 발견하기 위해서 업무와 업무간의 순서와 연관성을 찾았으며, 이러한 연관성을 찾기 위해 대량의 데이터에서 숨겨진 순서, 이상징후 처리, 선택적 프로세스 선정을 의사결정에 활용했다.

선택실행 프로세스, 동시실행 프로세스를 이용하고 중요한 프로세스이지만 수행 빈도가 낮아 발견되지 못하는 비즈니스 이상 프로세스를 발견하기 위해 데이터 마이닝의 순차패턴과 연관규칙을 이용하여 “A”사에 적합한 프로세스 마이닝 알고리즘을 제시하였다.

제시한 알고리즘은 실시간 모니터링 강화로 위험 인자를 조기에 발견하였고, 업무 혁신, 이상 징후의 실시간 조기 발견 및 빠른 대응으로 기업 이미지를 제고하는데 기여하였다.

본 논문에서는 기업의 영업 리스크 발생요인들을 상시 모니터링하고 예방 관리하기 위하여, 빅 데이터 환경 하에서 온라인 방식 및 전통적 감사에 최적으로 적용할 수 있는 프로세스 마이닝 적용방안을 제안하고 이를 평가하였다. 이 결과로 영업 감사에 대한 실시간 통합 모니터링이 강화되어 재무 리스크 회피, 감사기간 단축 및 감사 품질 개선 등의 효과가 나타났으며, 프로세스 마이닝이 감사분야에 적용된 기존 연구를 한층 더 강화 시켰다.

향후에는 자재 납기관리, 수요 및 생산 예측을 통한 계획생산과 생산성 관리체계 도입, 최적 조건 분석을 통한 효율적인 생산성 증대 및 관리에 초점을 두고자 한다.

참 고 문 헌 (Reference)

- [1] Agrawal, Divvyakant, P. Bernstein, E. Bertino, S. Davidson and Umeshwas, 2011.
- [2] Challenges and Opportunities with Big Data, Purdue e-Pubs, Cyber Center Technical Reports.
- [3] Barton, D. and D. Court, “Making Advanced Analytics Work For you,” Harvard Business Review, 79-83, October, 2012.
- [4] Biesdorf, S., D. Court and P. Wilnott, “Big data; What’s your plan,” The McKinsey Quarterly, March, 2013.
- [5] Diebold, Francis X., “Big Data’ Dynamic Factor Models for Macroeconomic Measurement and Forecasting,” Eighth World Congress of the Econometric Society, Cambridge University Press, 115-122, 2000.
- [6] Dumbil, E., Big Data Now: 2012 Edition, O’Reilly Media Inc., J-Pub Co, 2013.
- [7] Kart, L., High-Tech Tuesday Webinar: Big Data Opportunities in Vertical Industries, Gartner, August, 2012.
- [8] Kobielus, G. J., C. Moore, B. Hopkins and S. Coyne, Enterprise Hadoop: The Emerging Core of Big Data,

- October, Forrester, 2011.
- [9] Ericsson, Ericsson Mobility Report, February, 2015, available at: <http://www.ericsson.com/>
- [10] Research and Markets, Carrier B2B Data Revenue: Big Data, Analytics, Telecom APIs, and Data as a Service (DaaS) 2015-2020, July 2015, available at: <http://www.researchandmarkets.com/>
- [11] Analysys Mason, Big data analytics: how to generate revenue and customer loyalty using real-time network data, January 2013, available at: <http://www.analysismason.com/>
- [12] A. Cavoukian, Privacy by Design - The 7 Foundational Principles, available at: <https://www.privacybydesign.ca/>
- [13] Min Suk Song, Jae Yun Jung 'Process Mining Manifesto'

● 저 자 소 개 ●



유 영 석 (Young-Seok Yoo)

1992년 한국방송통신대학교 무역학과(학사)
2002년 아주대학교 경영대학원 경영학과(석사)
2014년~현재 한신대학교 대학원 정보통신학부(박사과정)
1983년~2001년 삼성물산(주) 관리부(부장)
2002년~2011년 (주)아이마켓코리아 영업총괄(상무)
2012년~현재 무림페이퍼(주) 상근감사
관심분야 : 내부감사, Process Mining, Big Data
E-mail : ys16.yoo@gmail.com



박 한 규 (Han-Gyu Park)

1994년 수원대학교 전자계산학과(학사)
1998년 수원대학교 대학원 전자계산학과(석사)
2015년~현재 한신대학교 대학원 정보통신학부(박사과정)
2007년~현재 한신대학교 IT계열 강사
관심분야 : 분산운영체제, 이동통신, 전자상거래
E-mail : guffer@hanmail.net



백 승 훈 (Seung-Hoon Back)

2015년 한신대학교 정보통신학과(학사)
2016년~현재 한신대학교 대학원 정보통신학부(박사과정)
2000년~2002년 폴더시스템(대표)
2015년~현재 (주)스마트크루(대표)
관심분야 : Big Data, 통계학
E-mail : hsfolder@nate.com



홍 성 찬 (Sung-Chan Hong)

1983년 고려대학교 통계학과(학사)
1990년 게이오대학 관리공학과(석사)
1994년 게이오대학 관리공학과(박사)
1994년~1995년 LG CNS 컨설팅사업부 책임연구원
1995년~1997년 상명대학교 정보처리학과 전임강사
2010년~2011년 한국인터넷정보학회 회장
1997년~현재 한신대학교 정보통신학부 교수
관심분야 : 정보시스템, 빅데이터
E-mail : schong@hs.ac.kr