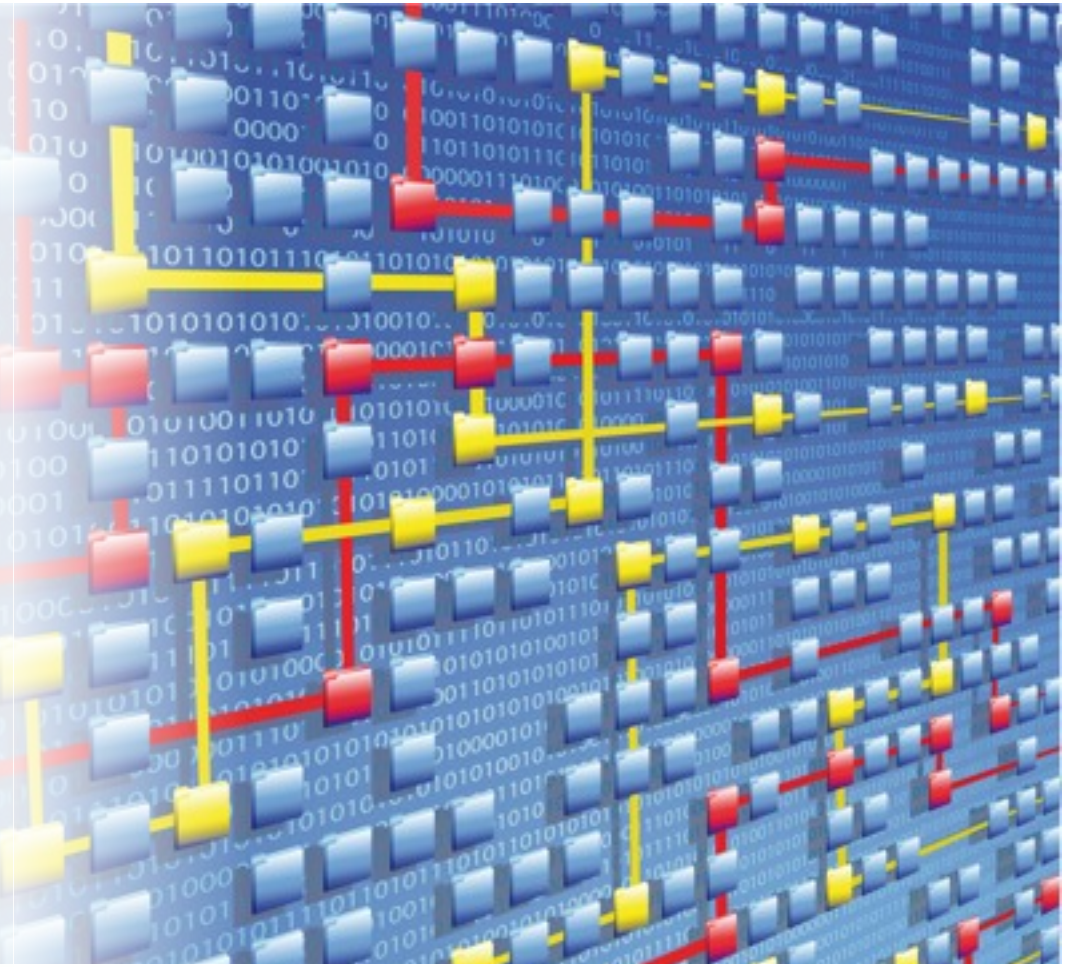


프로세스 마이닝 매니페스토



매니페스토는 어떤 단체의 “원칙과 목적을 나타낸 성명서”이다. 본 매니페스토는 **IEEE Task Force on Process Mining**의 회원과 지원자에 의해서 작성되었다. 본 태스크포스의 목표는 프로세스 마이닝에 대한 이해를 돕고, 프로세스 마이닝에 대한 연구 개발, 교육, 실제 적용 및 발전을 촉진하는 것을 목적으로 한다.

프로세스 마이닝은 비교적 새로운 연구 분야로, 계산 지능(**Computational Intelligence**) 및 데이터 마이닝과 프로세스 모델링 및 분석의 중간에 위치한 연구 분야이다. 프로세스 마이닝의 기본적인 아이디어는 정보시스템에 기록되어 있는 이벤트 로그에서 지식을 추출함으로써, 사람들이 머리 속에서 추정하고 있는 프로세스가 아닌 실제 업무 프로세스를 도출하고, 모니터링하며, 개선하는 것이다. 프로세스 마이닝은 (자동화된) 프로세스 도출(이벤트 로그에서 프로세스 모델 추출), 적합도 검사(모델과 이벤트 로그를 비교하여 차이점 분석), 소셜 네트워크/조직 마이닝, 시뮬레이션 모델의 자동 생성,

모델 확장, 모델 수정, 케이스 예측, 그리고 과거 데이터 기반 추천 등을 포함한다.

Contents:

| | |
|-----------------------------|----|
| 프로세스 마이닝 - State of the Art | 3 |
| 적용 가이드라인 | 6 |
| 도전과제 | 10 |
| 맺음말 | 14 |
| 용어사전 | 14 |

프로세스 마이닝 기술은 정보시스템에서 제공되는 이벤트 로그로부터 유용한 지식을 추출하는 연구로, 프로세스 도출(**discovery**), 모니터링(**monitoring**), 개선(**improvement**)을 위한 새로운 기법을 제공하며, 다양한 분야의 프로세스에 적용 가능하다. 프로세스 마이닝에 대한 관심은 분석 데이터의 가용성 측면과 프로세스 분석의 필요성 측면에서 설명할 수 있다. 데이터 가용성 측면에서 최근 트렌드를 보면, 정보시스템에서 더 많은 이벤트 로그 정보들을 수집하여, 프로세스의 히스토리에 대한 더 자세한 정보를 제공하고 있다. 또한 치열한 경쟁 속에서 급변하고 있는 기업 환경에서, 비즈니스 프로세스의 개선 및 지원에 대한 요구가 지속적으로 증대하고 있다. 이 두 가지 요인으로 인하여 프로세스 마이닝에 대한 관심이 점차 높아지고 있다. 본 매니페스토는 프로세스 마이닝 연구를 활성화하기 위해 **IEEE Task Force on Process Mining**에 의해 작성되었다. 본 매니페스토는 프로세스 마이닝을 위한 가이드라인과 중요한 도전과제들을 정리함으로써 소프트웨어 개발자, 연구자, 컨설턴트, 경영자, 사용자가 프로세스 마이닝을 이해하는 데 도움을 주고자 한다. 이를 통해서, 비즈니스 프로세스의 (재)설계, 관리, 지원 기능을 향상시킬 수 있는 새로운 도구로서, 프로세스 마이닝 기술의 성숙도를 높이는 데 기여하고자 한다.

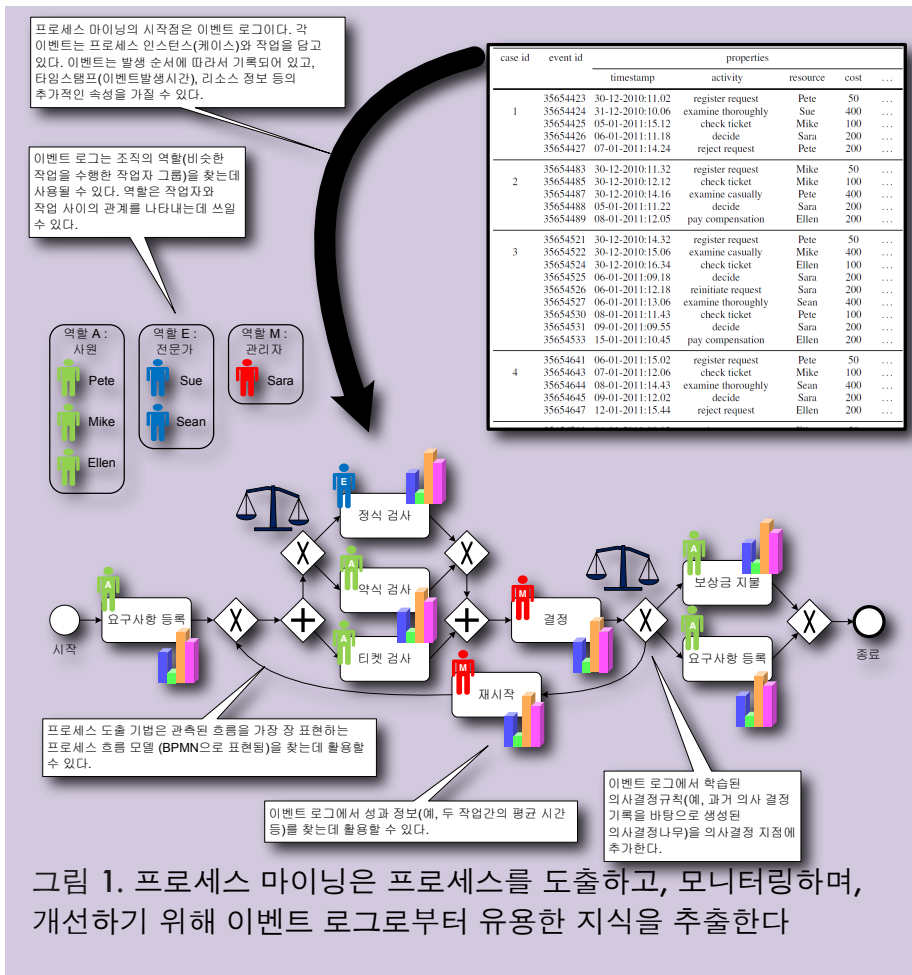


그림 1. 프로세스 마이닝은 프로세스를 도출하고, 모니터링하며, 개선하기 위해 이벤트 로그로부터 유용한 지식을 추출한다

프로세스 마이닝은 데이터 마이닝과 비즈니스 프로세스 모델링 및 분석의 중요한 연결 고리 역할을 한다. 프로세스 마이닝과 관련이 있는 기술로 비즈니스 인텔리전스(BI: Business Intelligence)가 있는데, 비즈니스 인텔리전스라는 명목하에 다소 단순한 리

포팅과 대시보드 도구를 지칭하는 여러 용어들이 소개되었다. 예를 들어 비즈니스 작업 모니터링(BAM: Business Activity Monitoring)은 비즈니스 프로세스의 실시간 모니터링을 가능하게 해주는 기술이다. 복합 이벤트 처리(CEP: Complex Event Processing)는 실시간으로 비즈니스를 모니터링하고 조정하며 최적화하기 위한 대용량 이벤트 데이터를 처리하는 기술을 의미한다. 경영 기법 관점에서 프로세스 마이닝과 관련된 기법으로 프로세스 또는 조직의 성과 측정을 강조하는 기업 성과 관리(CPM: Corporate Performance Management)가 있고, 이와 연관된 경영 기법으로 지속적 프로세스 혁신(CPI: Continuous Process Improvement), 비즈니스 프로세스 혁신(BPI: Business Process Improvement), 전사적 품질 경영(TQM: Total Quality Management), 그리고 식스 시그마(Six Sigma) 등이 있다. 이와 같은 경영 기법들은 프로세스 분석을 통해서 개선 가능성을 찾는다는 공통점이 있고, 이를 위해 프로세스 마이닝 기법이 효과적으로 활용될 수 있다.

BI 도구와 식스시그마, TQM 같은

경영 기법은 작업 시간 축소, 불량률 개선과 같은 운영 프로세스의 개선을 강조하고 있다. 이런 프로세스 개선 활동과 함께, 기업의 거버넌스, 리스크 관리, 규범 준수 등도 조직의 중요한 관심사가 되고 있다. 사베인스-옥슬리 법(Sarbanes-Oxley Act(SOX)), 바젤 II 협약(Basel II Accord)과 같은 규정은 규범 준수를 강조하고 있다. 프로세스 마이닝은 이런 규범 준수 여부를 보다 엄밀히 확인할 수 있고, 조직의 핵심 프로세스에 대한 정보의 유효성(validity)과 신뢰성(reliability)을 알아낼 수 있는 기법을 제공한다.

지난 10년 동안 이벤트 데이터의 확보가 용이해졌으며, 프로세스 마이닝 기술은 보다 발전하였다. 또한 앞서 언급한 바와 같이, 프로세스 개선(식스 시그마, TQM, CPI, CPM 등) 또는 규범 준수(SOX, BAM 등)와 관련된 경영 트랜드에 프로세스 마이닝을 활용할 수 있다. 프로세스 마이닝 알고리즘은 다양한 아카데미 시스템과 상용 시스템에서 구현되었다. 많은 연구자들이 프로세스 마이닝 분야에서 활발한 연구 활동을 하고 있으며, 프로세스 마이닝은 BPM(Business Process Management) 분야에서 가장 인기 있는 연구 토픽 중 하나이다. 프로세스 마이닝에 대한 기업의 관심도 높아지고 있으며, 다수의 소프트웨어 기업들이 자사의 제품에 프로세스 마이닝 기능을 포함시키고 있다. 예를 들면, 프로세스 마이닝 관련 기능이 있는 제품으로 ARIS Process Performance Manager (Software AG), Comprehend(Open Connect), Discovery Analyst (Stereo-LOGIC), Flow(Fourspark), Futura Reflect(Futura Process Intelligence), Interstage Automated Process Discovery(Fujitsu), OKT Process Mining suite(Exeura), Process Discovery Focus (Iontas/Verint), ProcessAnalyzer(QPR), ProM(TU/e) RbminerDbminer(UPC), Reflect|one(Pallas Athena) 등이 있다. 이와 같은 로그 기반 프로세스 분석에 대한 관심을 바탕으로 프로세스 마이닝 태스크포스를 설립하였다.

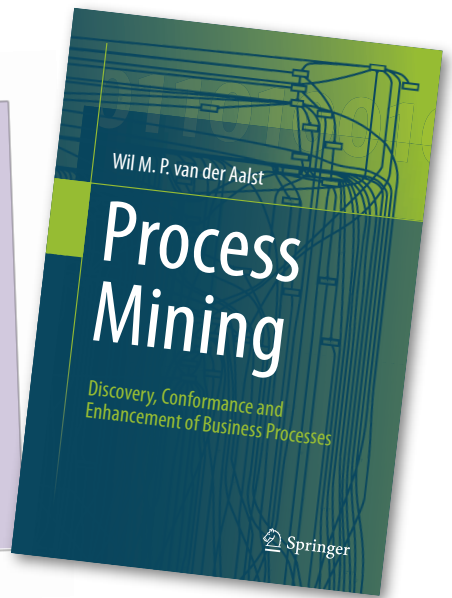
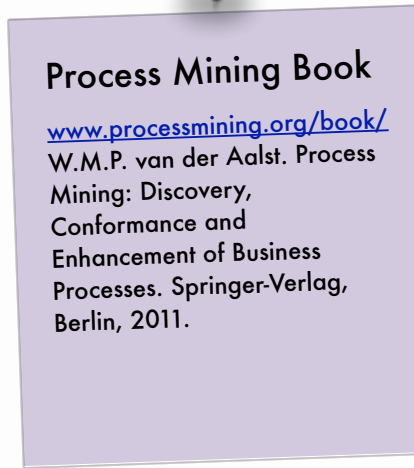
프로세스 마이닝 태스크포스는 Institute of Electrical and Electronics Engineers(IEEE)의 Computational Intelligence Society(CIS) 산하 Data Mining Technical Committee(DMTC) 소속으로 2009년에 구성되었다. 현재 태스크포스의 멤버는 소프트웨어 기업 (Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys,

프로세스 마이닝 태스크 포스의 조직 목적 :

- 1) 최종 수요자, 개발자, 컨설턴트, 경영자, 연구자들에게 최신 프로세스 마이닝 정보를 제공한다.
- 2) 프로세스 마이닝 기법과 도구의 이용을 촉진하며, 새로운 분야에 대한 적용을 장려한다.
- 3) 이벤트 데이터를 기록에 있어 표준화 작업을 한다.
- 4) 튜토리얼, 특별 세션, 워크샵, 패널을 조직한다.
- 5) 논문, 책, 비디오, 저널의 특별호를 출판한다.

Fluxicon, Businesscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, Sterologic), 컨설팅 회사 및 사용자(ProcessGold, Business Process-Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chili, BWI System GmbH, Excellentia, BPM, Rabobank), 연구기관(TU/e, University of Puda, Universitat Politècnica de Catalunya, New Mexico State University, Technical University of Lisbon, University of Calabria, Penn State University, University of Bari, Humboldt-Universität zu Berlin, Queensland University of Technology, Vienna University of Hafia, University of Bologna, Ulsan National Institute of Science and Technology, Cranfield University, K.U. Leuven, Tsinghua University, University of Innsbruck, University of Tartu) 등으로 구성되어 있다.

2009년 구성된 후, 위에 언급한 목적 달성을 위해 다양한 활동이 이루어졌다. 예를 들어, Business Process Intelligence 워크샵(BPI'09, BPI'10, BPI'11), IEEE 메인 컨퍼런스(CIDM'11)의 특별 트랙 같은 다양한 워크샵과 특별 트랙들이 태스크포스에 의해 준비가 되었다. 프로세스 마이닝을 소개 하기 위해 튜토리얼 세션(WCCI'10, PMPM'09 등)과 썸머스쿨(ESSCaSS'09, ACPN'10, CICH'10 등)을 준비하였으며, www.processmining.org 사이트를 통한 동영상 제공, 여러 논문 및 프로



세스 마이닝 서적 출판(Springer출판사) 활동이 이루어졌다. 또한 복잡한 대용량 이벤트 로그에서 유용한 정보를 추출해내는 비즈니스 프로세스 인텔리전스 챌린지(Business Process Intelligence Challenge(BPIC'11))를 개최하였다. 2010년 태스크포스는 표준 로그 포맷으로 XES(www.xes-standard.org)를 발표하였다. 이 포맷은 확장이 가능하며, OpenXES(www.openxes.org) 라이브러리와 ProM, XESame, Nitro 등의 도구에서 지원된다.

프로세스 마이닝 태스크포스의 활동에 대한 보다 자세한 정보는 <http://www.win.tue.nl/ieeetfpm/>를 참고하기 바란다.

2. 프로세스 마이닝: State of the Art

무어의 법칙은 정보시스템과 컴퓨팅에 기반을 두고 있는 관련 시스템들의 성능 향상의 특징을 잘 설명해 준다. 인텔의 공동창업자인 고든 무어는 1965년, 반도체칩에 저장되는 데이터의 양이 매년 두 배씩 늘어날 것이라고 예측하였다. 무어의 예측에 비해 약간은 더딘 페이스를 보였지만, 지난 50년간 성장은 폭발적이었다. 이러한 발전은 '디지털 세계'의 괄목할 만한 성장을 초래하여, 현재 모든 데이터는 전자적으로 저장되고 교환되고 있으며, 디지털 세계와 실제 세계는 더욱 동조화(align)되고 있다.

조직 내 프로세스와 매끄럽게 동조화 되어 있는 디지털 세계의 발전을 통해 이벤트를 기록하고 분석하는 것이 가능해졌다. 이벤트는 ATM에서의 현금 인출, 의사의 엑스레이 조작, 운전 면허 시험응시, 세금 신고, 여행자의 전자 항공권 수령 등 다양한 작업에 대해서 기록이 된다. 이런 이벤트 데이터를 프로세스에 대한 통찰 제공, 병목점 식별, 문제 예측, 규정 위반 내역 기록, 대책 권고, 프로세스 간소화 등을 위해 효과적으로 활용하는 것이 우리가 해결해야 할 과제이다.

프로세스 마이닝의 출발점은 이벤트 로그이다. 프로세스 마이닝 기법은 이벤트가 발생 순서대로 기록되어 있다는 것을 가정한다. 또한 하나의 이벤트는 프로세스의 단위 작업을 나타내며, 하나의 케이스(프로세스 인스턴스)와 관련을 맺고 있다. 이벤트 로그는 이벤트에 대한 부가적인 정보를 저장할 수 있다. 이런 부가적인 정보가 존재하면, 프로세스 마이닝 기법에 이를 활용할 수 있다. 예를 들어, 작업을 시작하거

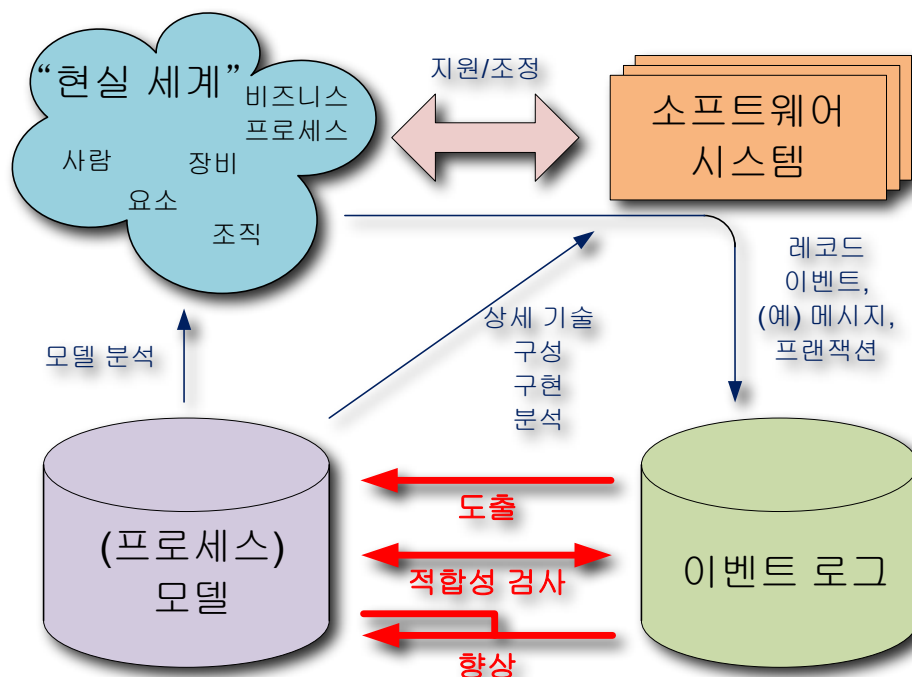


그림 2. 프로세스 마이닝 세 가지 주요 기법: (a) 프로세스 도출, (b) 적합도 검사 (c) 모델 향상.

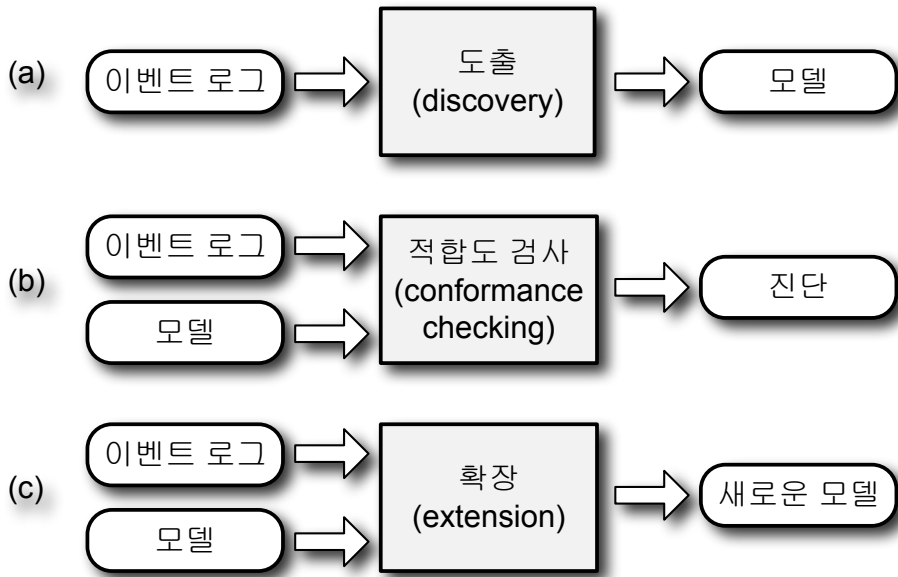


그림 3. 입력물과 산출물 관점에서 본 프로세스 마이닝 : (a) 프로세스 도출, (b) 적합도 검사 (c) 모델 향상

나 수행하는 리소스(업무수행자, 시스템, 장비 등), 이벤트의 타임스탬프(timestamp, 이벤트 발생 시간), 이벤트에 연관된 데이터(예, 주문의 규모) 등을 활용할 수 있다.

그림2와 같이, 이벤트 로그는 세 가지 유형의 프로세스 마이닝 수행에 이용될 수 있다. 프로세스 마이닝의 첫 번째 유형은 ‘도출(discovery)’이다. 도출 기법은 사전 정보 없이 이벤트 로그에서 모델을 생성하는 기법이다. 프로세스 모델 도출은 가장 중요한 프로세스 마이닝 기법이다. 많은 기관들이 현재 개발된 프로세스 마이닝 기법을 이용하여 이벤트 로그의 일부 케이스에서 실제 수행되고 있는 프로세스를 도출해 낼 수 있다는 것에 대해 놀라움을 느낀다. 프로세스 마이닝의 두 번째 유형은 ‘적합성(conformance)’이다. 적합성 검사는 보통 기존의 프로세스 모델과, 기존 모델에서 생성된 이벤트 로그를 비교한다. 적합성 검사는 로그에 기록된 현실이 모델에 일치하는지(또는 모델이 현실에 일치하는지) 확인하는데 이용할 수 있다. 적합성 검사는 프로세스 모델, 조직 모델, 선언적(declarative) 프로세스 모델, 비즈니스 규칙/정책, 법규 등 다양한 모델에 적용할 수 있다. 프로세스 마이닝의 세 번째 유형은 ‘향상(enhancement)’이다. 기본적인 아이디어는 이벤트 로그에 기록된 실제 프로세스 관련 정보를 이용하여 기존의 프로세스 모델을 확장하고 개선하는 것이다. 적합성 검사는 모델과 현실 사이의 정합성을 측정하는 반면에, ‘향상’은 이전 모델을 바꾸거나 확장시키는 것을 목표로 한다. 예를 들어, 이벤트 로그의 타임스탬프

를 이용하여 기존 모델을 병목점, 서비스 레벨, 업무 처리 시간, 빈도 등을 보여주는 모델로 확장할 수 있다.

그림3은 투입과 산출 관점에서 프로세스 마이닝의 세 가지 유형을 보여준다. 도출을 위한 기법들은 하나의 이벤트 로그를 입력받아 하나의 모델을 생성한다. 도출된 모델은 보통 Petri net, BPMN, EPC, UML activity diagram 등의 프로세스 모델이지만, 때로는 소셜 네트워크 같은 다른 관점의 모델이 될 수도 있다. 적합도 검사 기법은 이벤트 로그와 모델을 필요로 한다. 결과는 모델과 로그의 공통점과 차이점을 보여주는 진단 정보이다. 모델 향상(수정 또는 확장) 기법도 이벤트 로그와 모델을 필요로 한다. 결과는 개선된 모델 또는 확장된 모델이다.

프로세스 마이닝은 여러 관점을 다루고 있다. 프로세스 흐름 관점은 작업의 순서와 같은 프로세스의 흐름에 중점을 둔다. 이 관점의 목표는 모든 가능한 경로의 특징을 파악하는 것으로, 그 결과는 보통 Petri net 모델이나, EPC, BPMN, UML activity diagram과 같은 프로세스 모델로 표현된다. 조직 관점은 로그에 숨겨진 리소스에 대한 정보에 초점을 맞추어, 어떤 행위자(예, 작업자, 시스템, 역할, 부서)가 업무 수행에 관여하고, 그들이 어떻게 연결되어 있는지 분석한다. 조직 관점 분석의 목적은 역할이나 기능에 따라 사람들을 분류하여 조직 체계를 만들거나 사람들 사이의 소셜 네트워크를 보여주는 것이다. 케이스 관점은 케이스의 특징에 초점을 맞춘다. 하나의 케이스는 프로세스 내에서의 경로 또는 케이스에 참여하는 작업자를 통해 특성

프로세스 마이닝

특징:

1. 프로세스 마이닝은 프로세스 흐름의 도출에 한정되지 않는다.

이벤트 로그로부터 얻어진 프로세스 모델의 도출은 기업의 실무자와 학계의 연구자의 상상력에 불을 붙였다. 따라서 프로세스 흐름의 도출은 종종 프로세스 마이닝에서 가장 흥미진진한 부분으로 언급된다. 하지만 프로세스 마이닝은 프로세스 흐름의 도출에만 국한되어 있지 않다. 프로세스 도출은 프로세스 마이닝의 세 가지 유형(도출, 적합성, 향상) 중 하나에 불과하다. 또한 도출의 범위가 프로세스 흐름에 국한되어 있지 않고, 조직 관점과 케이스 관점, 시간 관점 또한 중요하다.

2. 프로세스 마이닝은 단순히 데이터 마이닝의 한 종류가 아니다.

프로세스 마이닝은 데이터 마이닝과 기존의 모델 기반 BPM 사이의 “잃어버린 연결 고리”라고 볼 수 있다. 대부분의 데이터 마이닝 기법은 프로세스 중심이 아니다. 병렬 구조를 내재한 프로세스 모델들은 의사결정나무나 연관규칙과 같은 단순한 데이터 마이닝의 구조로 표현할 수 없다. 따라서 완전히 새로운 형태의 표현법과 알고리즘이 필요하다.

3. 프로세스 마이닝은 오프라인 분석에 한정되지 않는다.

프로세스 마이닝 기법은 과거 이벤트 데이터로부터 정보를 추출한다. 사후 검토 데이터가 사용되고 있음에도 불구하고, 결과는 진행되고 있는 케이스에 적용될 수 있다. 예를 들어 도출된 프로세스 모델을 통해 부분적으로 완료된 고객의 주문을 완성하는 데 걸리는 시간을 예측할 수 있다.

을 분석할 수 있다. 또는 케이스와 연관된 데이터의 값에 따라서 케이스의 특성을 분석할 수도 있다. 예를 들어, 보충 주문을 나타내는 케이스가 있으면, 주문된 상품의 수량과 공급자는 유용한 정보가 된다. 시간 관점은 이벤트의 시간과 빈도를 고려한다. 이벤트가 타임스탬프를 포함하면, 병목점 도출, 서비스의 레벨 측정, 리소스 활용도 모니터링, 수행 중인 케이스의 잔여 시간 예측 등이 가능하다.

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 생각한다. 이에 대한 설명은 이전 페이지 “프로세스 마이닝 특징”을 참조하기 바란다.

그림4는 BPM(Business Process Management) 라이프사이클을 통해 프로세스 마이닝의 위치를 보여준다. BPM 라이프사이클은 비즈니스 프로세스와 관련 정보시스템의 7단계를 보여준다. “(재)설계” 단계에서 새로운 프로세스 모델이 생성되거나 기존의 프로세스 모델이 수정된다. “분석” 단계에서는 후보 모델과 대체 모델들에 대한 분석이 이루어진다. “(재)설계” 단계를 거치고 나면 모델이 구현되거나 (“구현” 단계), 혹은 기존의 시스템이 “(재)구성” 된다. “수행” 단계에서는 설계된 모델이 수행이 되며, 프로세스가 모니터링된다. 또한 사소한 수정 사항들은 재설계 단계를 거치지 않고 수정된다 (“조정” 단계). “진단” 단계에는 실행된 프로세스가 분석되고, 분석 결과를 통해 얻은 정보를 바탕으로

새로운 프로세스를 재설계 하기도 한다. 프로세스 마이닝은 그림4의 대부분의 모든 단계에서 유용하게 사용된다. 특히 “진단” 단계는 프로세스 마이닝 기법의 사용이 매우 유용하게 활용된다. 하지만 그 활용이 “진단” 단계에만 국한되지 않는다. 예를 들면, “실행” 단계에서 프로세스 마이닝의 기법들을 운영적 지원을 위해 활용할 수 있다. 과거 데이터를 기반으로 학습된 모델을 바탕으로 예측과 추천을 통해 수행 중인 케이스에 영향을 줄 수 있다. 유사한 형태의 의사 결정이 프로세스를 수정하거나 프로세스의 “(재)구성”에 활용할 수 있다.

그림5는 프로세스 마이닝 활동과 결과물에 초점을 맞춰, 프로세스 마이닝 프로젝트에 나올 수 있는 단계들에 대해서 보여준다. 프로세스 마이닝 프로젝트는 계획 작성과 계획에 대한 타당성 기술로 시작한다 (0단계). 프로젝트를 시작한 후에는 시스템, 도메인 전문가, 경영진으로부터 이벤트 데이터와 모델, 목적, 의문사항 등을 도출한다 (1단계). 이 단계에서는 가용 데이터(분석을 하려면 어떤 자료가 사용될 수 있을까?), 도메인(어떤 것이 중요한 질문인가?), 그림5의 1단계의 중간 산출물(예를 들어, 과거 데이터, 핸드메이드 모델, 목적, 의문점 등)에 대한 이해가 요구된다. 2단계에는 자동화된 프로세스 도출 기법을 활용하여 프로세스 흐름 모델을 도출하고, 이벤트 로그에 연결한다. 도출된 프로세스 흐름 모델은 의문 사항에 대한 답을 제시할 수 있고, 이를 바탕으로 기존 프로세스를 재설계하거나 수정할 수 있다. 또한

적용 가이드라인:

GP1: 이벤트 데이터를 일등석 손님처럼 다루어야 한다

GP2: 로그 추출은 명확한 질문을 바탕으로 이루어져야 한다

GP3: 동시작업, 선택을 포함한 기본적인 프로세스 흐름 모델 생성자를 지원해야 한다

GP4: 이벤트는 모델 엘리먼트와 연결되어 있어야 한다

GP5: 모델은 실제 상황을 목적에 맞게 추상화해야 한다

GP6: 프로세스 마이닝은 하나의 연속적인 프로세스이다

모델을 활용하여 이벤트 로그를 필터링하거나 수정할 수 있다. 예를 들어 빈도수가 낮은 작업이나 이상치에 해당하는 케이스를 제거하거나, 누락된 이벤트를 추가할 수 있다. 때로는 동일한 프로세스 인스턴스에 속하는 이벤트를 서로 연관시키는데 상당한 노력을 기울여야 한다. 로그에 있는 이벤트는 프로세스 모델의 개체에 연결되어야 한다. 프로세스가 비교적 구조적일 때는 3단계에서 프로세스 모델을, 데이터, 시간, 리소스 관점으로 확장할 수 있다. 2단계에서 생성된 이벤트 로그와 모델의 관계를 바탕으로 모델을 확장할 수 있다. 예를 들어 이벤트와 연관된 타임스탬프를 통해 각 작업의 대기 시간을 예측할 수 있다. 이렇게 확장함으로써 부가적인 질문에 대한 답을 찾을 수 있고, 추가적인 분석을 시작할 수 있다. 궁극적으로 3단계에서 형성된 모델들은 운영적 지원의 목적으로 사용할 수 있다(4단계). 과거 이벤트 데이터에서 추출된 지식은 수행 중인 케이스의 정보와 결합되고, 이는 수행 중인 프로세스에 대한 개입, 예측, 추천에 사용된다. 3단계와 4단계는 프로세스가 충분히 안정적이고 체계적일 때 도달할 수 있다.

현재, 그림5의 모든 단계를 지원할 수 있는 기법과 도구가 존재하기는 하지만, 프로세스 마이닝은 상대적으로 새로운 패러다임이기 때문에 이용 가능한 도구가 미흡한 실정이다. 또한 전

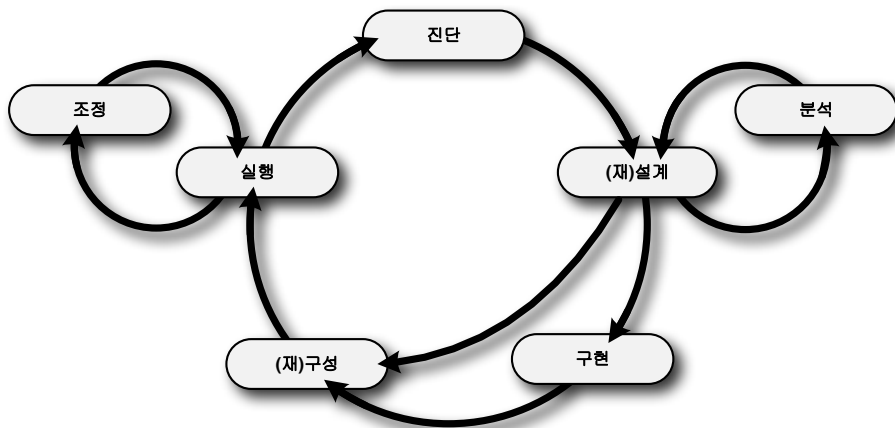


그림 4. 비즈니스 프로세스와 관련 정보시스템의 BPM 라이프사이클 7 단계: 프로세스 마이닝은 “구현” 단계를 제외한 모든 단계에 활용할 수 있다.

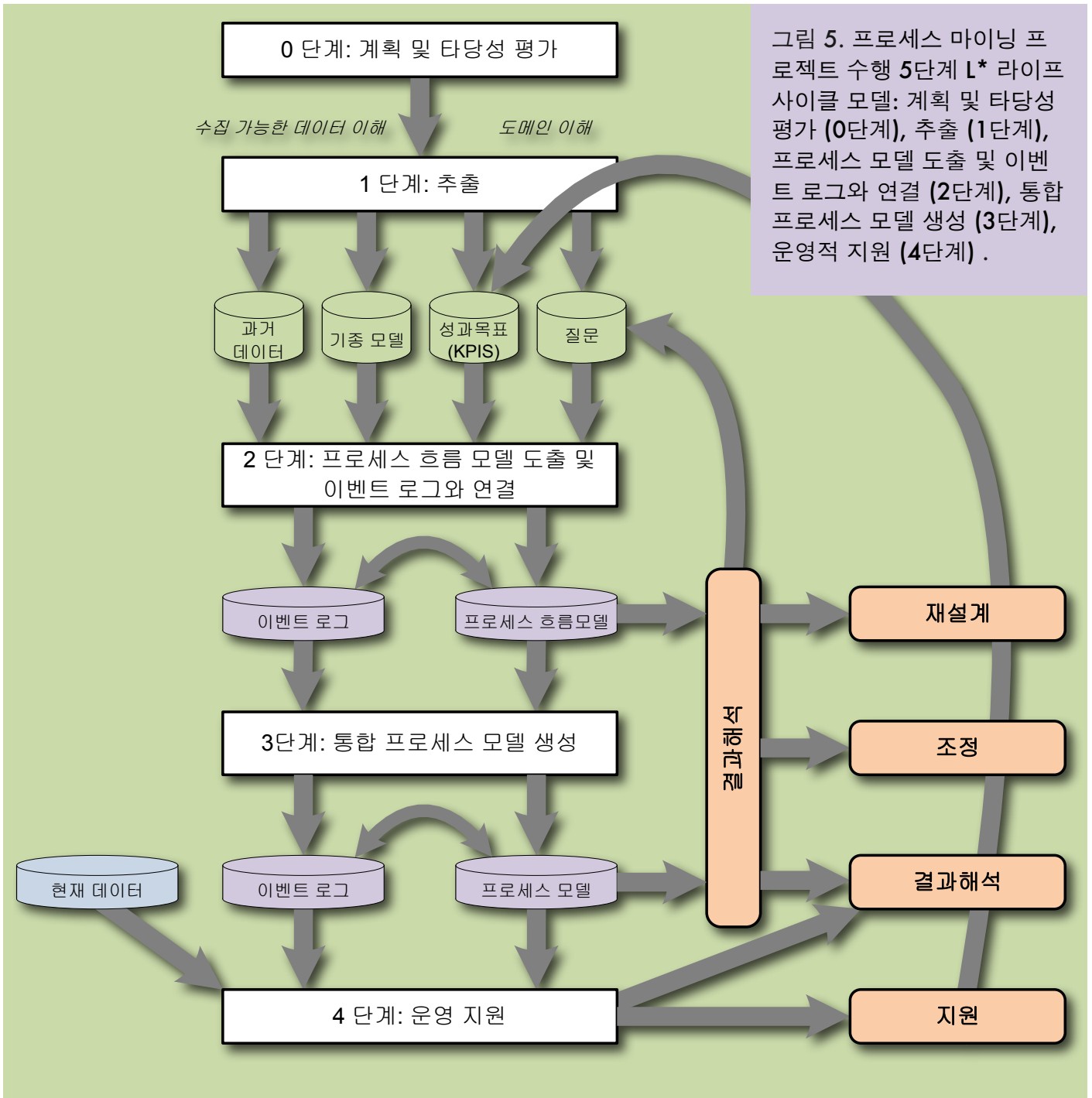


그림 5. 프로세스 마이닝 프로젝트 수행 5단계 L* 라이프 사이클 모델: 계획 및 타당성 평가 (0단계), 추출 (1단계), 프로세스 모델 도출 및 이벤트 로그와 연결 (2단계), 통합 프로세스 모델 생성 (3단계), 운영적 지원 (4단계).

향적인 사용자들은 종종 프로세스 마이닝의 잠재성과 한계성을 인식하지 못하고 있다. 따라서 본 문서는 프로세스 마이닝 기법의 사용자와 프로세스 마이닝 최신 기술을 개발하는데 관심이 있는 연구자와 개발자를 위해 프로세스 마이닝의 적용 가이드라인(3장)과 도전과제(4장)를 제공한다.

3. 적용 가이드라인

모든 새로운 기술이 그랬던 것처럼, 프로세스 마이닝 또한 실제 상황에 적용할 때 여러 가지 실수를 범할 수 있다. 따라서 본 장에서는 사용자나 프

로세스 분석가가 프로세스 마이닝 기술을 적용하는 데 있어서 실수를 줄일 수 있는 6가지 가이드라인을 제시한다.

GP1: 이벤트 데이터를 일등석 손님처럼 다루어야 한다

프로세스 마이닝의 시작점은 이벤트이다. 이벤트의 집합을 이벤트 로그라고 하며, 이벤트가 반드시 특정 이벤트 로그 파일의 형태로 저장되어 있을 필요는 없다. 이벤트는 데이터베이스 테이블, 메시지 로그, 메일 보관함, 트랜잭션 로그, 또는 그 외 다양한 데이

터 형태로 저장될 수 있다. 데이터의 형태보다 더 중요한 것은 이벤트 로그의 품질이다. 프로세스 마이닝의 결과는 이벤트 로그에 가장 큰 영향을 받는다. 따라서 이벤트 로그는 분석 대상 프로세스를 지원하는 정보시스템에서 가장 중요한 자원이다. 하지만, 불행히도 많은 경우 이벤트 로그는 단순히 디버깅이나 프로파일링을 위한 부산물로 취급된다. 예를 들어, 필립스(Philips) 헬스케어의 의료 장비는 이벤트를 기록하는데, 이는 특별한 의도가 있는 것이 아니라 단순히 소프트웨어 개발자가 "프린트 명령문"을 넣어 두었기 때문이다. 물론 장비 개발에 있어, "프린

| 레벨 | 특징 | 예시 |
|-------|--|--|
| ★★★★★ | 최상 레벨: 이벤트 로그의 품질이 최상이며(신뢰성(trustworthy)과 완벽성(complete) 확보), 이벤트가 잘 정의되어 있다. 이벤트는 자동적이며(automatic), 체계적이며(systematic), 믿을 수 있고 (reliable), 안정적인 방법(safe manner)으로 기록된다. 프라이버시와 보안 이슈들이 적절하게 다루어지고, 기록된 이벤트(모든 속성값 포함)가 명확한 의미를 가지고 있다. 이는 하나 또는 둘 이상의 온톨로지가 정의되어 있고, 이벤트와 이벤트의 속성값이 온톨로지와 연결되어 있음을 의미한다. | 의미에 대한 주석이 달린 BPM 시스템의 로그 |
| ★★★★ | 이벤트들이 자동적이고, 체계적이며, 믿을 수 있는 방법으로 기록이 되어 있다. 로그는 신뢰할 수 있고(trustworthy), 완벽하다(complete). 3단계(★★★)에 있는 시스템과는 달리, 프로세스 인스턴스(case)와 단위 작업(activity)이 명확하게 표현되어 있다 | 일반적인 BPM/workflow 시스템에서 생성되는 로그 |
| ★★★ | 이벤트가 자동으로 기록되지만, 로깅 방법이 체계적이지 않다. 하지만, 2단계(★★) 로그와는 달리, 이벤트 로그가 현실을 어느 정도 반영하고 있다. (이벤트 로그에 신뢰성은 있으나, 반드시 완벽하지 않아 모든 이벤트가 기록되지 않는다.) 예를 들어, ERP 시스템에서 기록되는 로그가 3단계에 해당된다. ERP 로그는 여러 테이블에서 추출하는 과정이 필요한데, 로그의 신뢰성은 있으나 로깅 방법이 체계적이지 않다. | ERP 시스템의 테이블, CRM 시스템의 이벤트 로그, 메시징 시스템의 트랜잭션 로그, 하이테크 시스템의 이벤트 로그 (예, 의료 장비) 등 |
| ★★ | 이벤트가 자동으로 기록이 된다. 즉, 정보시스템의 부산물로 이벤트가 기록된다. 커버리지가 다양하다. 즉, 어느 이벤트를 기록해야 하는지에 대한 체계적인 방법이 없다. 또한 정보시스템을 우회할 수도 있어서, 누락이 되거나 적절히 기록되지 않는 이벤트가 있다. | 문서 관리 시스템의 로그, 제품 관리 시스템의 로그, 임베디드 시스템의 로그, 서비스 엔지니어의 작업 기록 등 |
| ★ | 최하 단계: 이벤트 로그의 품질이 매우 낮다. 기록된 이벤트가 현실을 제대로 반영하지 못하고 이벤트가 누락된 경우가 있다. 보통 이벤트 로그가 수작업으로 기록된다. | 종이 기반 문서 결재 기록, 종이 기반 진료 기록 등 |

표 1. 이벤트 로그의 성숙도 레벨

트 명령문"을 넣는 비공식적인 가이드 라인이 있기는 하지만, 이보다는 이벤트 로그의 품질을 높이기 위하여 보다 체계적인 로깅 방법이 필요하다. 즉, 이벤트 데이터를 일등석 손님으로 인식되어야 한다.

이벤트 데이터의 품질을 판단하기 위해 다양한 기준이 있다. 먼저, 이벤트는 신뢰성이 높아야 한다(trustworthy). 즉, 기록된 이벤트는 반드시 일어난 일이고, 이벤트의 속성값은 정확해야 한다. 이벤트 로그는 완전해야 한다(complete). 주어진 범위에 있어서, 누락되는 것이 없어야 한다. 또한, 기록된 이벤트의 의미가 잘 정의되어 있어야 한다. 이벤트 데이터를 기록할 때 프라이버시나 보안과 관련된 이슈들을 명확히 다루어야 한다. 예를 들

어, 이벤트에 관련된 업무 수행자는 관련 이벤트가 기록된다는 것을 인지하고 있어야 하며, 향후 어떤 분석에 사용될는지 알고 있어야 한다.

표1은 이벤트 로그의 성숙 레벨을 최상(★★★★★)에서 최하(★)까지 5점 척도로 나타내고 있다. 예를 들어, 필립스 헬스케어의 이벤트 로그는 ★★★ 레벨에 해당이 된다. 이벤트가 자동으로 기록되며, 기록된 정보는 정확하다. 하지만, 이벤트의 의미 정의와 이벤트의 범위 정의가 체계적이지 않다. 프로세스 마이닝은 ★★★ 레벨 이상의 로그에만 적용이 가능하다. 원칙적으로 프로세스 마이닝을 ★ 레벨과 ★★ 레벨의 로그에 적용하는 것이 가능하지만, 로그의 분석에 있어 많은 문제가 발생할 수 있고, 분석 결과에 대한 신뢰성이

낮아지게 된다. 사실, ★ 레벨의 로그에 프로세스 마이닝을 적용하는 것은 무의미하다. 프로세스 마이닝으로 의미 있는 결과를 얻기 위해서는 가능한 상위 레벨의 로그를 대상으로 분석을 수행해야 한다.

GP2: 로그 추출은 명확한 질문을 바탕으로 이루어져야 한다

그림5에서 볼 수 있듯이, 프로세스 마이닝 작업은 질문을 기반으로 진행되어야 한다. 명확한 질문(분석 요구 사항)없이, 의미있는 이벤트 데이터를 추출하는 것은 매우 힘들다. 예를 들어, SAP과 같은 ERP 시스템에 수천 개의 테이블이 존재하는데, 명확한 질문 없이 관련된 테이블을 선택하여 데이

터를 추출하는 것은 불가능하다.

그림1과 같은 프로세스 모델은 특정 타입의 케이스의 라이프사이클을 나타낸다. 따라서 프로세스 마이닝 기법을 적용하기 전에 분석의 대상이 되는 케이스를 선택할 필요가 있다. 케이스의 선택은 분석 요구 사항에 따라 결정하는 것이 중요하다. 고객 주문의 처리를 예로 들어보자. 고객은 한 번의 주문으로 여러 제품을 주문할 수 있기 때문에 각 고객 주문은 여러 주문 라인으로 구성될 수 있다. 하나의 고객 주문은 여러 번의 배송으로 처리될 수 있으며, 또한 하나의 배송은 여러 주문의 주문 라인을 처리할 수 있다. 따라서, 주문과 배송은 다대다 관계를 가지고, 주문과 주문 라인은 일대다 관계를 가진다. 주문, 주문 라인, 배송에 관련된 이벤트 데이터를 포함한 데이터베이스의 경우, 다양한 프로세스 모델을 도출할 수 있다. 개별 배송에 대한 라이프사이클을 찾기 위한 목적으로 프로세스 모델을 도출할 수 있고, 각각의 주문 라인, 또는 각 배송의 라이프사이클을 찾기 위하여 프로세스 모델을 도출할 수도 있다.

GP3: 동시작업, 선택을 포함한 기본적인 프로세스 흐름 모델 생성자를 지원해야 한다

BPMN, EPC, Petri net, BPEL, UML activity diagram 등 여러 종류의 프로세스 모델링 언어가 존재한다. 이 중 일부는 다양한 모델링 엘리먼트들을 지원한다. 예를 들어 BPMN의 경우, 50가지가 넘는 그래픽 엘리먼트들을 지원한다. 일부 모델링 언어들은 단순한 엘리먼트들만 지원한다. 예를 들어 Petri net의 경우 플레시스, 트랜지션, 아크의 세 가지 요소로만 구성된다. 플로우 구조는 프로세스 모델에서 중추적인 역할을 한다. 모든 언어에서 지원되는 기본적인 워크플로우 구조는 순차(sequence), 병렬(AND분기/병합), 선택(XOR분기/병합), 순환(Loop)을 지원한다. 프로세스 마이닝 기술은 이런 기본적인 패턴을 지원한다. 그러나 일부 프로세스 마이닝 기술은 병렬 관계를 지원하지 못하고, 마코브 체인과 트랜지션 시스템만 지원한다.

그림6은 병렬(AND분기/병합) 관계를 도출할 수 없는 프로세스 마이닝 기법을 사용한 경우를 보여 준다. 이번

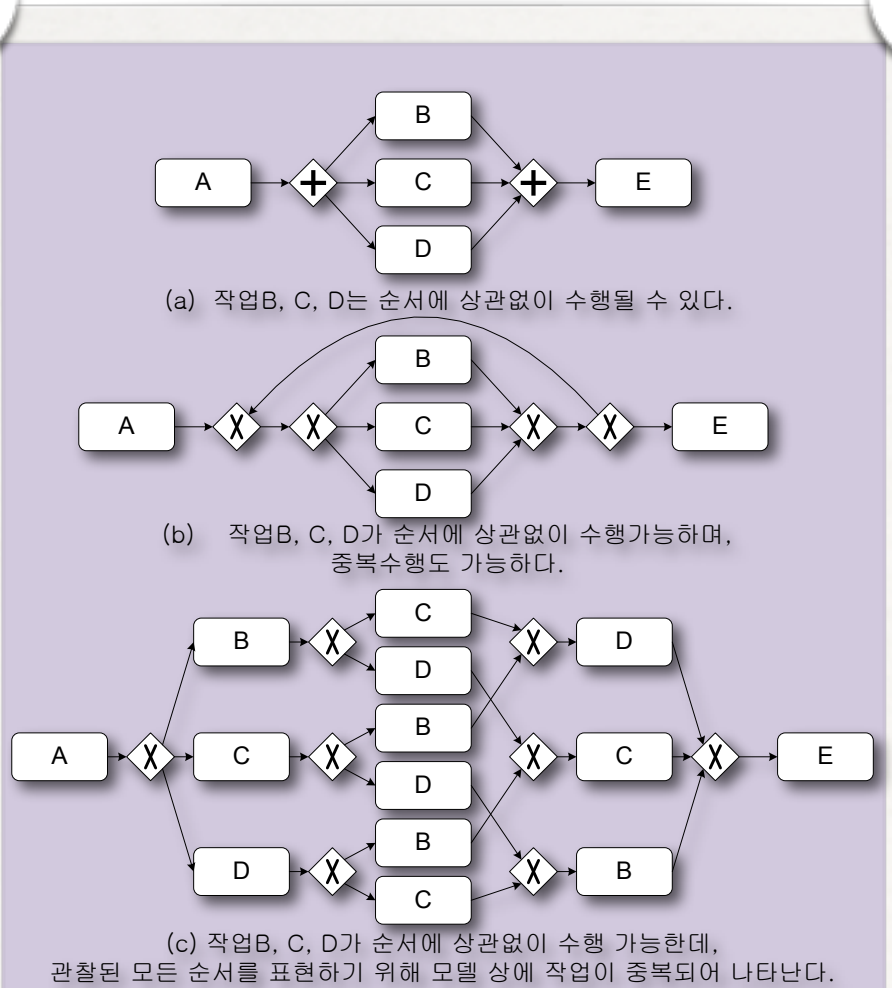


그림 6. AND 분기와 병합과 같은 동시작업을 프로세스 도출 알고리즘이 표현할 수 없는 경우에 발생하는 문제를 보여준다. 예제에서 작업B, 작업C, 작업D는 병렬 수행된다. 만약 병렬 수행 가능한 10개의 모델이 있다면 1024 (2¹⁰)가지 상태가 존재하며, 3,628,800 (10!)개의 가능한 수행 가능 패턴이 나타난다.

트 로그L이 {<A,B,C,D,E>, <A,B,D,C,E>, <A,C,B,D,E>, <A,C,D,B,E>, <A,D,B,C,E>, <A,D,C,B,E>}라고 가정할 경우, 로그L에 속한 모든 케이스는 작업A로 시작을 하고 작업E로 끝나게 된다. 작업B, 작업C, 작업D는 작업A와 작업E 사이에 어디든지 올 수 있다. 그림6(a)의 BPMN 모델은 두 개의 AND 게이트웨이를 사용하여, 로그L의 프로세스 모델을 표현한 것이다. 만약 사용되는 프로세스 마이닝 기법이 AND 게이트웨이를 찾을 수 없다면, 그림6(b)와 그림6(c) 같은 모델이 생성될 수 있다. 그림6(b)의 BPMN 모델은 모델 자체는 간략하지만, 로그L에서 나타나는 패턴 이외에 더 많은 케이스 패턴을 생성할 수 있다. 예를 들어, 프로세스 모델에서 로그L에 없는 <A, B, B, B, E>와 같은 패턴도 생성할 수 있다. 그림6(c)의 BPMN 모델은 로그L에 나오는 패턴만

표현이 되지만, 자세히 보면 로그L의 패턴을 모두 표현하여 모델링했기 때문에 프로세스를 간략하게 표현하지 못하고 있다. 이 예제는 병렬 수행을 포함하는 실제 프로세스 모델을 대상으로 프로세스 마이닝을 적용하는 경우에, 프로세스 마이닝 기법이 병렬 작업을 제대로 처리하지 못하면, 과소적합 모델 생성(실제 케이스 패턴 이외에 여러 다른 패턴을 포함) 또는 아주 복잡한 모델 생성을 할 수 있는 문제를 보여준다.

그림6에서 보여 주었듯이, 프로세스 마이닝 기법이 기본적인 워크플로우의 패턴을 지원하는 것이 중요하다. 위에서 언급한 기본적인 패턴 이외에, 포괄적 선택 구조(inclusive decision)와 부분 동기화(partial synchronization)를 간단히 표현할 수 있도록 OR 분기와 병합을 지원하는 것이 바람직하다.

GP4: 이벤트는 모델 엘리먼트와 연결되어 있어야 한다

2장에서 언급하였듯이, 프로세스 마이닝이 프로세스 흐름 모델을 도출하는 것만 가능하다는 것은 잘못된 생각이다. 그림1에서 보는 바와 같이, 도출된 프로세스 모델은 조직 관점, 시간 관점, 데이터 관점 등 다양한 관점을 보여준다. 또한 그림3과 같이, 모델 도출은 프로세스 마이닝의 세 가지 유형 중 하나에 불과하다. 다른 두 가지 유형의 프로세스 마이닝은 모델의 엘리먼트와 로그의 이벤트 간의 관계에 영향을 받는다. 이 관계는 모델에 이벤트 로그를 재현(replay)하는 데 사용할 수 있다. 재현은 이벤트 로그와 모델의 불일치를 파악하는데 사용할 수 있다. 예를 들어 로그의 특정 이벤트들은 모델로부터 나타날 수 없는 경우가 있다. 프로세스 마이닝의 적합도 검사 기법은 이런 불일치의 계량화 및 진단에 이용할 수 있다. 이벤트 로그의 타임스탬프는 재현 중에 시간에 따른 행동 패턴을 분석하는 데 사용할 수 있다.

어떤 경우에는 이런 관계를 도출하는 것이 쉽지 않다. 예를 들어 하나의 이벤트가 두 개의 작업과 연결이 된 경우나, 어떤 작업을 지칭하는 지 불분명한 경우에는 관계 도출이 쉽지 않은데, 모호한 경우는 프로세스 마이닝 결과의 정확한 해석을 위해 제거해야 한다. 이벤트와 작업 사이의 문제뿐 아니라 이벤트와 프로세스 인스턴스 사이의 모호성 문제도 발생할 수 있다. 이런 이벤트와 작업 사이의 관계, 이벤트와 프로세스 인스턴스 사이의 관계를 이벤트 상관 관계라고 지칭한다.

GP5: 모델은 실제 상황을 목적에 맞게 추상화해야 한다.

이벤트 데이터에서 도출된 모델은 실제 상황에 대한 뷰를 제공한다. 이런 뷰는 이벤트 로그로부터 수집된 행동 패턴을 추상화하여 나타낸다. 주어진 이벤트 로그에서 다양한 뷰를 생성할 수 있고, 사용자에게 따라서 서로 다른 뷰를 필요로 할 수도 있다. 즉, 이벤트 로그에서 도출되는 모델은 방향을 찾기 위한 지도와 같이 프로세스를 분석하기 위한 지도라고 볼 수 있다. 본 장

에서는 이런 모델을 사용하는 데 있어 중요한 두 가지 사항을 설명한다.

첫째로, 특정 지역에 대한 유일한 지도가 존재하지 않는다는 점을 명심해야 한다. 목적에 따라서 도로 지도, 하이킹 지도, 자전거 지도 등 여러 종류의 지도가 존재한다. 이런 여러 지도는 모두 같은 지역을 보여 주고 있는데, 만약 모든 정보가 통합된 하나의 지도를 만드려고 하면, 어떤 목적에도 적합하지 않은 지도가 만들어지게 될 것이다. 프로세스 모델도 이와 마찬가지로 지이다. 프로세스 모델에 나타나는 정보는 사용자의 요구에 맞추어 사용자가 원하는 정보를 강조하여 표현해야 한다. 도출된 모델은 프로세스 흐름, 데이터, 시간, 리소스, 비용 등 여러 관점 중 필요한 부분에 초점을 맞추고 있으며, 서로 다른 입자 수준과 정확도로 정보를 제공해야 한다. 예를 들어, 경영자의 경우는 비용에 대한 정보에 초점을 맞춘 개략적인 프로세스 모델을 원하는 반면, 프로세스 분석가는 정상적 흐름에서 벗어난 변이를 찾기 위한 보다 상세한 모델을 원할 수 있다. 또한 이해관계자에 따라서 서로 다른 레벨의 정보를 제공해야 한다. 전략적인 레벨의 경우, 장기적 관점에서 결정을 내리기 위해, 오랜 기간의 이벤트 정보를 취합하여 분석 정보를 제공해야 한다. 전술적인 레벨의 경우, 중기적 관점에서 최근의 데이터 경향을 제공해야 하고, 운영 레벨의 경우, 즉각적인 판단을 위해 현재 수행중인 케이스의 데이터를 활용해야 한다.

둘째로, 이해하기 쉬운 모델을 만들기 위해, 지도 제작(cartography)에 사용되는 아이디어를 활용할 필요가 있다. 예를 들어, 도로 지도의 경우, 중요하지 않은 도로나 도시는 추상화하여 간략하게 나타낸다. (예를 들어, 도시 내의 도로나 도시 주변 지역을 하나의 도시로 표현할 수도 있다.) 또한 중요하지 않은 정보는 삭제할 수 있고, 중요한 정보를 강조하기 위해 색깔을 사용할 수 있다. 또한 지도에 사용되는 기호의 크기로 중요성을 표현하기도 한다. 예를 들어 선의 굵기와 점의 크기로 중요성을 표현할 수 있다. 또한 지도는 가로축(동서방향)과 세로축(남북방향)에 대한 명확한 해석이 가능하다. 따라서 지도 상의 지형물의 X좌표와 Y좌표는 명확한 해석이 가능하다.

이렇게 지도에서 사용되는 색깔, 크기, 위치 정보들은 우리가 일반적으로 사용하는 프로세스 모델에서는 찾아 볼 수 없다. 그러나 이런 요소들을 프로세스 마이닝에서 생성되는 프로세스 모델에 쉽게 적용할 수 있다. 예를 들어, 작업의 빈도수나 작업의 중요도(비용, 리소스 사용)에 따라서 프로세스 모델에 표현되는 작업의 크기를 다르게 할 수 있을 것이다. 또한 두 작업의 상관관계에 따라서 작업을 연결하는 연결선의 두께를 다르게 할 수 있고, 연결선의 색깔을 통해 병목점을 표현할 수 있다.

이처럼 사용자에게 따라서 적합한 표현 방법을 선택하고, 정제된 정보를 제공하는 것이 중요하다. 이는 사용자에게 프로세스 마이닝 결과를 가시화하고, 적합한 모델을 도출하기 위한 프로세스 마이닝 알고리즘을 개발하는 데 중요하다 (도전과제 C5 참조).

GP6: 프로세스 마이닝은 하나의 연속적인 프로세스이다

프로세스 마이닝은 이벤트 데이터에 직접 연결되는 의미있는 “지도(도출된 모델)”를 제공한다. 과거의 데이터와 현재의 데이터는 도출된 모델에 투영될 수 있다. 또한 프로세스 분석 중에 분석 대상 프로세스가 변경될 수도 있다. 이러한 프로세스의 동적인 특성으로 인해, 프로세스 마이닝은 단순히 한번 수행하는 작업으로 보는 것은 바람직하지 않다. 프로세스 마이닝의 목적은 단순히 하나의 고정된 모델을 도출하는 것이 아니라, 프로세스 모델에 생명력을 불어 넣어 프로세스 수행자와 분석가가 매일매일 프로세스의 진행 상태를 관찰할 수 있도록 한다.

이런 특성을 지도에 다양한 정보를 태깅하는 매시업(mashup)과 비교할 수 있다. 예를 들어 구글 지도를 활용하여 교통 정보, 부동산 정보, 패스트푸드점, 영화관 등의 정보를 지도에 투영한 수천 가지의 매시업 서비스가 존재한다. 사람들은 이런 서비스 상에서 줌인/줌아웃 등의 기능을 활용하여 필요한 정보를 얻기 위한 효과적으로 상호작용한다. 예를 들어 교통 혼잡 정보가 지도에 투영되고, 사용자는 상세한 정보를 보기 위하여 혼잡 지역을 선택하

여 추가 정보를 살펴볼 수 있다. 프로세스 마이닝도 실시간 이벤트 데이터를 바탕으로 분석을 수행할 수 있다. 지도에서처럼 이벤트가 GPS 좌표를 가지고 실시간으로 지도에 투영이 되는 것을 생각해 볼 수 있다. 자동차 내비게이션 시스템처럼 프로세스 마이닝 도구는 사용자가 프로세스를 내비게이션할 수 있고, 프로세스 모델에 병목점과 같은 동적인 정보를 투영할 수 있으며, 수행되는 케이스의 예상 종료 시간과 같은 예측 정보를 제공할 수 있다. 즉 프로세스 모델은 보다 동적으로 활용될 수 있으며, 따라서 프로세스 마이닝은 업무 수행에 도움이 되는 동적인 정보를 다양한 스케일(분/시간/일/주/달)로 제공하는 연속적인 프로세스로 보아야 한다.

4. 도전과제

프로세스 마이닝은 단순하지 않은 운영 프로세스를 관리해야 하는 최근 조직에게 매우 유용한 도구이다. 엄청난 양의 이벤트 데이터가 증가하고 있는 동시에, 규범 준수, 효율성 제고, 고객 서비스 향상 등 다양한 요구를 만족시키기 위하여, 프로세스와 정보를 완벽히 동조화할 필요가 있다. 프로세스 마이닝은 다양한 분야에서 활용될 수 있지만, 여전히 중요한 도전과제들이 남아있으며, 이는 프로세스 마이닝이 새롭게 등장한 분야라는 것을 의미한다. 이 장에서는 프로세스 마이닝의 대표적인 도전과제들을 소개한다. 여기에 프로세스 마이닝의 모든 도전과제를 제공하고 있는 것은 아니다. 시간이 지나면서 새로운 도전과제가 등장할 수 있고, 프로세스 마이닝의 발전과 함께 일부 과제들은 해결될 수도 있을 것이다.

C1: 이벤트 데이터의 발견, 병합, 정제

프로세스 마이닝을 위해 적절한 이벤트 데이터를 추출하는 데에는 여전히 많은 노력이 소요된다. 이 때 극복해야 하는 대표적인 난관들은 다음과 같다.

- 데이터가 다양한 소스에 “분산”되어 있을 수 있다. 그러면, 이 정보들을 병합할 필요가 있다. 만약 서로 다른 데이터 소스에서 상이한 아이디어를 사용한다면 문제가 될 수 있다. 예를 들어, 한 시스템에서는 사람들을 식

별하기 위하여 이름과 생년월일을 사용하는데, 다른 시스템에서는 주민등록번호를 사용하는 경우가 있다.

- 이벤트 데이터들은 종종 “프로세스 중심”이 아니라 “객체 중심”으로 기록된다. 예를 들어, 개별 제품, 팔레트, 컨테이너가 RFID 태그를 가지고 있어서, 이벤트 기록들이 이 태그를 참조하고 있을 수 있다. 특정 고객의 주문을 분석하기 위해서는 이와 같은 객체 중심의 이벤트들을 병합할 필요가 있다.
- 이벤트 데이터가 불완전할 수도 있다. 흔히 일어나는 문제로 이벤트들이 프로세스 인스턴스를 명시적으로 가리키고 있지 않다는 것이다. 일반적으로 이 정보를 도출하는 것이 가능하지만 많은 노력이 소요된다. 그리고 일부 이벤트에는 시간 정보가 누락되었을 수도 있다. 필요한 시간 정보를 얻기 위하여 타임스탬프를 대략적으로 추정해야 하는 경우도 있다.
- 이벤트 로그가 이상치(outlier)를 포함할 수 있다. 예를 들어, 예외적 행위 또한 노이즈(noise)로 간주될 수 있다. 이벤트 데이터를 정제하기 위해서는 “이상치를 어떻게 정의할 것인가?” “이상치를 어떻게 탐지할 것인가?” 같은 문제들을 해결해야 한다.
- 로그가 서로 다른 입자 수준의 이벤트를 포함하고 있을 수 있다. 병원 정보시스템의 이벤트 로그에 단순한 혈액 검사에서부터 복잡한 수술 과정까지 함께 포함하고 있을 수 있다. 타임스탬프 또한 밀리센컨드(28-9-2011:h11m28s32ms342)부터 일자(28-9-2011)까지 다양한 수준의 정보를 가지고 있을 수 있다.
- 이벤트는 보통 특정한 상황(context) (날씨, 작업부하, 요일 등) 하에서 발생한다. 그 상황은 그 때의 특별한 현상을 이해하는 데 도움을 줄 수 있다. 예를 들면, 처리 중인 작업의 개수가 많거나 휴일에는 평소보다 응답 시간이 더 길어질 수 있다. 분석을 위해서는 이러한 상황 정보를 결합하는 것이 바람직하다. 즉, 이벤트 데이터를 상황 데이터와 병합해야 한다는 의미이다. 이 때 너무 많은 변수를 추가하면 분석 처리가 힘들어져서 “차원의 저주(curse of dimensionality)”에 빠질 수도 있다.

도전과제:

- C1: 이벤트 데이터의 발견, 병합, 정제
- C2: 다양한 특징을 가지는 복합 이벤트 로그 다루기
- C3: 대표적인 벤치마크 창출
- C4: 개념 변화 다루기
- C5: 프로세스 도출에서 표현적 편중 개선
- C6: 품질 기준 간의 균형: 적합도, 간결성, 정확도, 일반성
- C7: 다조직 마이닝
- C8: 운영적 지원 제공
- C9: 다른 유형의 분석과 프로세스 마이닝의 결합
- C10: 비전문가들을 위한 사용자 개선
- C11: 비전문가들을 위한 이해도 개선

위와 같은 문제를 해결하기 위해서는 더 나은 도구와 방법론이 필요하다. 앞서 지적한 바와 같이, 조직은 이벤트 로그를 부산물이 아니라 일등석 손님으로 다루어야 한다. 그 목표는 ★★★★★ 레벨의 이벤트 로그를 얻는 것이다(표1 참고). 고품질 이벤트 로그를 얻기 위해서는 데이터웨어하우스 분야의 기법들을 참조할 필요가 있다. 예를 들어, 데이터 입력시 수행하는 간단한 확인 만으로도 이벤트 데이터 오류를 엄청나게 줄이는 데 도움이 될 것이다.

C2: 다양한 특징을 가지는 복합 이벤트 로그 다루기

이벤트 로그는 다양한 특징을 가지고 있을 수 있다. 어떤 이벤트 로그는 너무 방대하여 다루기 힘들 수도 있고, 어떤 이벤트 로그는 너무 작아서 의미

있는 결론을 얻기에 데이터가 충분하지 않을 수도 있다.

일부 도메인에서는 놀랄 만큼 엄청난 양의 이벤트가 기록된다. 그래서, 성능이나 확장성을 개선하기 위해 부가적인 노력이 필요하다. 예를 들어, ASML은 삼성이나 텍사스 인스트루먼트와 같은 여러 조직에서 사용되고 있는 모든 웨이퍼 스캐너를 항상 모니터링하고 있다. (현재 약 70%의 칩이 ASML의 웨이퍼 스캐너를 이용하여 생산되고 있다.) 기존 도구들은 이러한 도메인에서 수집되는 페타바이트 (petabyte) 수준의 데이터를 처리할 수 없다. 기록되는 이벤트 양 이외에도, 케이스 당 기록되는 이벤트의 평균 개수, 케이스들의 유사성, 서로 다른 이벤트의 개수, 서로 다른 경로의 개수 등과 같이 다양한 특성들이 존재한다. 예를 들어, 이벤트 L1이 1000개 케이스를 가지고 있고, 케이스 당 평균 10개의 이벤트를 가지고 있으며, 케이스들이 서로 거의 차이가 없다고 가정해 보자. (즉, 여러 케이스들이 거의 동일하거나 거의 유사한 경로를 따른다.) 이벤트 L2는 100개의 케이스를 가지고 있지만, 케이스 당 평균 100개의 이벤트를 가지고 있으며, 모든 케이스들이 서로 다른 경로를 따른다고 가정해 보자. 두 이벤트 로그는 유사한 규모(약 10,000 개의 이벤트)를 가지고 있음에도 불구하고, 분명히 L1보다는 L2가 분석하기 훨씬 어렵다.

이벤트 로그가 동일한 행위를 포함하고 있다고 해서, 그것은 완전하다고 가정해서는 안된다. 프로세스 마이닝 기법들은 “개방된 세상을 가정(open world assumption)”하고 있기 때문에 불완전성을 다룰 필요가 있다. 즉, 발생하지 않았다고 해서 발생할 수 없다는 것을 의미하는 것은 아니다. 이는 큰 변동성을 가진 소규모의 이벤트 로그에 대한 도전과제를 제시한다.

앞서 언급한 바와 같이, 어떤 로그는 매우 낮은 추상화 수준의 이벤트를 포함하고 있다. 이러한 로그는 너무 방대하여 담당자들이 이처럼 낮은 수준의 개별 이벤트에는 관심이 없는 경우가 많다. 그러므로 하위 수준의 이벤트는 상위 수준의 이벤트로 결합하는 것이 좋다. 예를 들어, 특정 그룹의 환자에 대한 진료 및 치료 프로세스를 분

석할 때, 병원 실험실의 정보시스템에 기록된 개별 시험에 대해서는 관심이 없을 지도 모른다. 이 때, 이벤트 로그가 프로세스 마이닝에 적합한 지 살펴보기 위하여, 조직은 여러 번의 시행착오가 필요할 수 있다. 이를 위해 특정 데이터 집합이 주어졌을 때, 프로세스 마이닝 도구는 신속한 분석 타당성 테스트(feasibility test)를 수행할 수 있어야 한다. 이 테스트는 잠재적인 성능 문제를 알려주고, 너무 불완전하거나 너무 상세한 로그에 대해서 경고를 해 줄 수 있어야 한다.

C3: 대표적인 벤치마크 창출

프로세스 마이닝은 비교적 최근에 등장한 기술이다. 이는 훌륭한 벤치마크가 아직 존재하지 않음을 의미한다. 예를 들어, 수십 가지의 프로세스 도출 기법들을 활용할 수 있고, 다양한 벤더들이 여러 가지 제품들을 제공하고 있지만, 이러한 기법들의 품질에 관한 합의가 없다. 기능과 성능 측면에서 서로 엄청난 차이가 있음에도 불구하고, 다양한 기법과 도구들을 비교하기가 쉽지 않다. 그러므로, 예제 데이터 집합과 대표적인 품질 기준으로 구성된 훌륭한 벤치마크를 개발할 필요가 있다.

전통적인 데이터 마이닝 기법에 관해서는 많은 훌륭한 벤치마크가 존재한다. 이러한 벤치마크는 분석 도구 제공자 및 연구자들에게 자신들이 개발한 기법의 성능을 개선하도록 자극해 왔다. 프로세스 마이닝의 경우에 이는 더 큰 과제이다. 예를 들어, 1969년에 Codd가 소개한 관계형 모델은 매우 단순하여 널리 지원을 받았다. 그 결과 하나의 데이터베이스로부터 다른 데이터베이스로 데이터를 이전하는 데 거의 노력이 들지 않고, 해석할 필요가 없다. 프로세스의 경우, 아직 이와 같은 단순한 모델이 없다. 프로세스 모델링으로 제시된 표준들은 훨씬 더 복잡하여, 동일한 개념들을 정확히 지원하는 벤더가 거의 없다. 이는 프로세스가 관계형 데이터보다 훨씬 더 복잡하기 때문이다.

그럼에도 불구하고, 프로세스 마이닝을 위한 대표적인 벤치마크를 창출하는 것은 매우 중요한 일이다. 일부 초기 작업이 진행되었다. 예를 들면, 프로세스 마이닝 결과의 품질을 측정

하는 다양한 평가지표로서 적합도 (fitness), 간결성(simplicity), 정확도 (precision), 일반성(generalization)이 제시된 바 있다. 나아가, 몇 가지 이벤트 로그가 공개적으로 이용 가능하다 (cf. www.processmining.org). 예를 들어, 본 태스크 포스가 조직한 제 1회 비즈니스 프로세스 인텔리전스 챌린지 (BPIC'11)에서 사용된 이벤트 로그가 있다 (cf. doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc54).

한편으로는 실제 데이터 집합에 기반한 벤치마크가 있어야 하고, 다른 한편으로는 특정한 성격을 반영하는 가상 데이터 집합을 생성할 필요가 있다. 이러한 가상 데이터 집합은 불완전한 이벤트 로그, 노이즈가 포함된 이벤트 로그, 특수한 형태의 프로세스 집단 등을 분석하기 위한 프로세스 마이닝 기법을 개발하는 데 도움이 될 것이다.

대표적인 벤치마크 창출과 더불어, 프로세스 마이닝 결과의 품질(C6도 참고)을 판정하는 데 사용될 기준에 관해서도 합의가 필요하다. 나아가, 그 결과를 판정하는 데 데이터 마이닝의 교차검증 기법을 적용할 수 있다. k-fold 교차검증(cross validation)을 예로 들 수 있다. 이벤트 로그를 k 부분으로 분할한 후, k-1 부분은 프로세스 모델을 학습하는 데 사용하고, 나머지 한 부분은 그 결과를 판정하는 데 사용할 수 있다. 이를 k회 반복하여 모델의 품질을 개선할 수 있다.

C4: 개념 변화 다루기

개념 변화(concept drift)라는 용어는 분석하는 동안 프로세스가 변화하는 상황을 의미한다. 예를 들어, 이벤트 로그의 초기 부분에서는 두 개의 작업의 병렬 수행이 가능하였지만, 이벤트 로그의 후반부에는 순차적 실행으로 변화하였을 수도 있다. 주기적인 또는 계절적인 변화에 의하여 프로세스가 변화할 수도 있으며(예를 들면, “12월에 수요가 많다”거나, “금요일 오후에는 가능한 직원이 거의 없다”), 환경 변화에 의하여 프로세스가 변화할 수도 있다(예를 들면, “시장의 경쟁이 더 심화되고 있다”). 그러한 변화는 프로세스에 영향을 미치고, 우리는 그 변화를 탐지하고 분석하는 것이 매우 중요하다. 프로세스의 개념 변화는 이벤트

로그를 더 작은 로그로 분할하여 “발자취(footprints)”를 분석함으로써 도출할 수 있다. 이후에 훨씬 더 큰 이벤트 데이터에 대한 “2차적 분석”이 필요하다. 그럼에도 불구하고, 안정된 상태에 존재하는 프로세스는 거의 없으므로, 개념 변화를 이해하는 것은 프로세스 관리에 있어서 매우 중요한 일이다. 그러므로 개념 변화를 적절히 분석할 수 있는 더 많은 연구와 도구 지원이 필요하다.

C5: 프로세스 도출에서 표현적 편중 개선

프로세스 도출 기법은 특정 언어(예. BPMN, Petri net 등)를 사용하여 모델을 생성한다. 그러나 실제 프로세스를 도출할 때 사용된 프로세스 언어와 결과의 가시화를 분리하는 것이 중요하다. 대상 프로세스 언어를 선택하는 것은 종종 암묵적인 가정을 수반하기 때문이다. 프로세스 도출은 검색 공간(search space)에 한정된다. 즉, 선택된 언어로 표현할 수 없는 프로세스는 도출할 수가 없다. 프로세스 도출 과정에서 발생하는 이런 “표현적 편중(representational bias)”은 프로세스 도출 기법을 선택할 때 고려되어야 하며, 단지 그래픽적 표현을 선호한다고 해서 특정 기법을 선택해서는 안 된다.

그림6의 예제를 살펴보면, 목표 언어가 병렬 작업을 표현할 수 있는 지는 도출한 모델을 가시화하는 데 영향을 줄 뿐만 아니라, 그 알고리즘이 고려하고 있는 모델의 유형에도 영향을 준다. 만약 표현적 편중으로 인해 병렬수행을 표현할 수 없고(그림6(a)가 표현 불가능), 동일한 이름을 여러 작업에 사용할 수 없다면(그림6(c)가 표현 불가능), 그림6(b)에 제시된 모형과 같이 잘못된 모형만을 도출할 수 있을 것이다. 이는 표현적 편중에 대하여 얼마나 신중하고 치밀하게 선택할 필요가 있는지 보여주는 예제이다.

C6: 품질 기준 간의 균형: 적합도, 간결성, 정확도, 일반성

이벤트 로그는 불완전한 경우가 많다. 즉, 모든 행위를 포함하고 있는 것이 아니라, 일부 예제만 주어진다. 프로세스 모델은 보통 기하급수적으로 많은 개수의 다른 흔적(trace)을 포함하거나, 또는 (Loop의 경우에는) 무한

개의 다른 흔적을 가질 수도 있다. 나아가, 일부 흔적은 다른 흔적들보다 훨씬 낮은 확률을 가질 수도 있다. 그러므로, 이벤트 로그에 가능한 흔적들이 모두 포함되어 있다고 가정하는 것은 비현실적이다. 로그가 완전하다고 가정하는 것이 비현실적이라는 것을 보여주는 예제를 살펴보자. 병렬로 수행 가능한 10개의 작업으로 구성된 프로세스가 있고, 해당 로그가 약 10,000개의 케이스 정보를 포함하고 있다고 가정하자. 10개의 병렬 작업이 가질 수 있는 중간 과정은 총 $10! = 3,628,800$ 가지나 된다. 즉, 잠재적인 흔적(3,628,800개)에 비하여 훨씬 작은 케이스(10,000개)가 존재하기 때문에, 이 로그 안에 모든 중간 과정이 포함되어 있을 수 없다. 로그에 수백만 개의 케이스가 존재한다고 하더라도, 모든 가능한 변형들이 존재하기는 거의 어렵다. 문제를 복잡하게 하는 또 다른 이유는 일부 흔적은 다른 것들보다 발생 빈도가 적다는 것이다. 이들이 “노이즈”로 간주될 지도 모른다. 이러한 노이즈 행위를 포함시켜 합리적인 모형을 만드는 것은 거의 불가능하다. 도출한 모델은 이들로부터 추상화시킬 필요가 있으며, 즉, 적합도 검사를 통하여 빈도가 낮은 행위들은 조사하는 것이 더 바람직하다.

노이즈와 불완전성으로 인하여 프로세스 도출은 여전히 도전적인 과제이다. 실제로, (a) 적합도(fitness), (b) 간결성(simplicity), (c) 정확도(precision), (d) 일반성(generalization)와 같은 네 가지 경쟁적인 품질 차원이 존재한다. 적합한 모델이란 로그에 나타난 대부분의 행위들을 표현할 수 있어야 한다. 한 모델이 로그의 모든 흔적들을 시작부터 끝까지 모두 재현할 수 있다면, 그 모델은 완벽한 적합도를 가지는 모델이다. 로그에 나타난 행위를 표현할 수 있는 모델 중에서는 가장 간결한 모델이 가장 훌륭한 모델이다. 이 원리는 오컴의 면도날(Occam's Razor)이라고 알려져 있다. 도출한 프로세스 모델의 품질을 판정하는 데 있어서 적합도와 간결성 중 하나만으로는 충분하지 않다. 예를 들어, 이벤트 로그 내의 모든 흔적들을 재현할 수 있는, 극히 간결한 Petri net 모델(“flower model”)은 쉽게 생성할 수 있지만, 그 모델은 동일한 작업들로 표현되는 어떤 다른 이벤트 로그들도 포함하게 된다. 마찬가지로, 단지 이벤트 로그에

표현된 행위들만 정확히 표현하는 모델도 바람직하지 않다. 로그는 일부 예제 행위만을 포함하고 있을 뿐, 발생 가능한 많은 흔적들이 아직 나타나지 않았을 수 있다는 것을 기억하자. 정확한 모델이란 “ 지나치게 많은 ” 행위를 포함하지 않는 모델이다. 분명히, “flower model”은 정확성이 부족하다. 정확하지 않은 모델은 “과소적합(underfitting)”된 모델이다. 과소적합이란 모델이 로그 내의 예제 행위들을 지나치게 일반화하고 있다는 것이다. (즉, 모델이 로그에 나타난 것들과는 너무 다른 행위들도 표현하고 있다는 것이다.) 모델은 로그에 나타난 예제만으로 행위를 국한하지 말고 일반화시켜야 한다. 일반화되지 않은 모델은 “과적합(overfitting)”된 모델이다. 과적합이란 로그는 일부 예제 행위들만 포함하고 있는 데에도 불구하고, 너무 특화된 모델을 생성한 것이다. (즉, 그 모델은 특정 예제 로그만을 표현하고 있고, 동일 프로세스의 또 다른 예제 로그로부터 완전히 다른 프로세스 모델이 생성되는 경우이다.)

적합도, 간결성, 정확도, 일반성의 균형을 맞추는 것은 도전과제이다. 이 때문에 대부분의 강력한 프로세스 도출 기법들이 다양한 인자를 제공하고 있다. 경쟁적인 네 가지 품질 차원의 균형을 잘 맞추기 위하여 개선된 알고리즘을 개발할 필요가 있다. 나아가, 모든 인자들은 최종 사용자가 이해하기 쉬워야 한다.

C7: 다조직 마이닝

일반적으로 프로세스 마이닝은 단일 기업 내에 적용된다. 그러나, 서비스 기술, 공급사슬 통합, 클라우드 컴퓨팅이 더욱 확산되면서, 다수 조직의 이벤트 로그에 대한 분석이 가능한 시나리오들이 있다. 다조직 프로세스 마이닝(cross-organizational process mining)을 필요로 하는 경우는 주로 두 가지 형태로 구분할 수 있다.

첫 번째로는 프로세스 인스턴스를 처리하기 위하여 서로 다른 조직들이 함께 협력하는 협력적 환경을 고려해 보자. 이러한 다조직 프로세스는 “직소 퍼즐(jigsaw puzzle)”처럼 간주할 수 있다. 즉, 케이스를 성공적으로 완료하기 위하여 전체 프로세스가 부분으로 쪼개어져서 협력 조직들에 분산되어 있다. 이 중 한 조직만의 이벤트 로그를 분석하는 것은 의미가 없다. 종단간

프로세스를 도출하기 위해서는 여러 조직의 이벤트 로그들을 병합해야 한다. 이러한 작업은 조직의 경계를 넘어서 관련 이벤트들을 연결해야 하므로 단순한 작업이 아니다.

다음으로는 경험, 지식, 공동 인프라를 공유하면서 완전히 동일한 프로세스를 수행하는 다조직을 고려할 수 있다. **Salesforce.com**과 같은 예를 살펴보자. **Salesforce.com**은 여러 조직의 영업 프로세스를 관리하고 지원한다. 이 조직들은 프로세스, 데이터베이스 등의 인프라를 공유하고 있는 반면에, 동일한 프로세스의 여러 변형들을 지원하기 위해 시스템의 형상 변경할 수 있기 때문에 엄격한 프로세스 모델을 따르도록 강요하지 않는다. 또 다른 예로 지방정부에서 수행하는 기초적인 프로세스들(예. 건축허가 승인)을 고려해보자. 국가의 모든 지방정부들이 동일한 기본 프로세스를 지원할 필요는 있지만, 또한 차이가 있을 수 있다. 서로 다른 조직 간의 변형들을 분석하는 것은 분명히 흥미로운 일이다. 다조직 프로세스 마이닝 결과에 기초하여 각 조직들은 서로에게 배울 수 있고, 서비스 제공자들은 자신의 서비스를 개선하여 부가가치 서비스를 제공할 수 있다. 다조직 프로세스 마이닝의 두 가지 형태 모두를 위하여 현재의 분석 기법들을 개발할 필요가 있다. 이 기법들은 비밀과 보안 문제 또한 고려해야만 한다. 조직은 경쟁적 이윤이나 신뢰 부족으로 인하여 정보 공유를 원하지 않을 수 있다. 그러므로 비밀 유지에 입각한 프로세스 마이닝 기법을 개발하는 것도 중요한 일이다.

C8: 운영적 지원 제공

초기에 프로세스 마이닝은 과거 데이터 분석에 중점을 두었다. 그러나 최근에는 많은 데이터 소스들이 거의 실시간으로 업데이트되고 있으며, 이벤트들이 발생할 때 분석할 수 있는 충분한 계산 능력이 지원되고 있다. 그러므로 프로세스 마이닝을 오프라인 분석으로 국한해서는 안되며, 온라인 운영적 지원(**operational support**)에 사용될 수 있어야 한다. 탐지(**detect**), 예측(**prediction**), 추천(**recommend**)의 세 가지 운영적 지원이 가능하다. 케이스가 미리 정의된 프로세스로부터 벗어나는 순간 이를 감지하여 시스템이 알림을 보내줄 수 있다. 보통 이러한 통보는 (실제로 영향을 줄 수 있도록) 즉

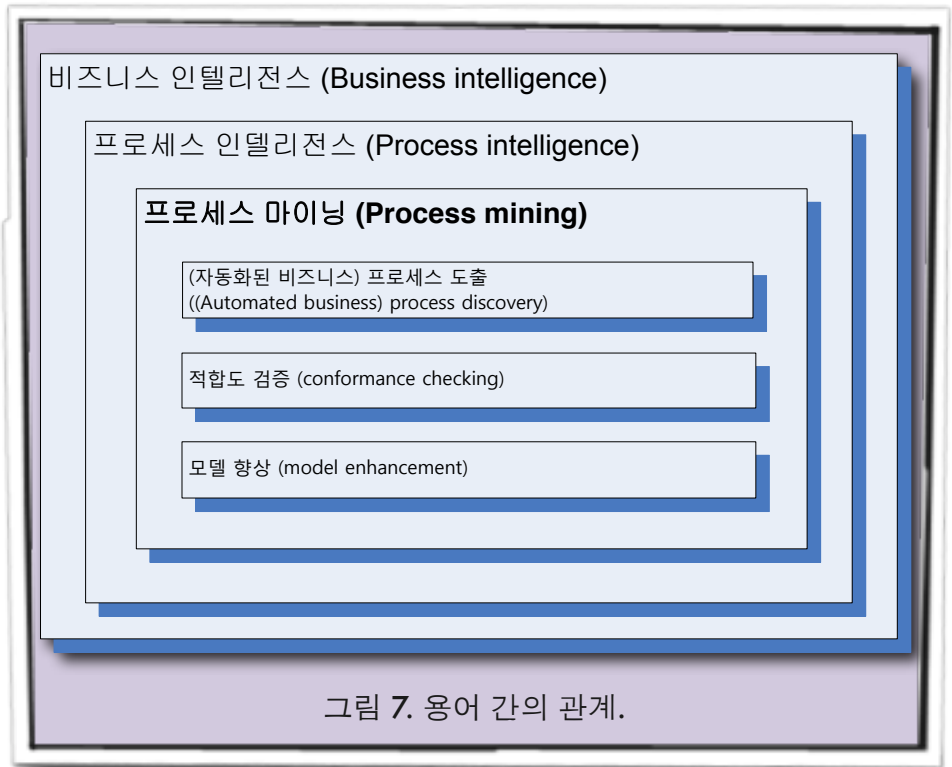


그림 7. 용어 간의 관계.

시 생성하는 것이 좋으며, 오프라인 방식으로는 부적합하다. 예측 모델을 구축하기 위해서 과거 데이터를 사용할 수 있다. 이는 실행 중인 프로세스 인스턴스를 안내하는 데 사용할 수 있다. 예를 들어, 한 케이스의 남은 종료 시간을 예측하는 것이 가능하다. 또한 이 예측에 근거하여 비용을 줄이거나 흐름 시간을 단축하기 위하여 특정한 활동을 제안하는 추천 시스템을 구축할 수 있다. 프로세스 마이닝 기법을 온라인으로 적용하는 것은 계산 능력과 데이터 품질 측면에서 또 하나의 도전과제라고 할 수 있다.

C9: 다른 유형의 분석과 프로세스 마이닝의 결합

운영 관리 (**operations management**)와 오퍼레이션 리서치 (**operations research**)는 모델링을 이용하는 경영 과학 (**management science**) 분야이다. 여기에는 선형계획법과 프로젝트 계획으로부터 대기 모형, 마코프 체인, 시뮬레이션에 이르기까지 다양한 수리적 모델이 사용된다. 한편, 데이터 마이닝은 “예상치 못한 관계를 찾거나 데이터를 요약하기 위하여, 데이터 소유자가 이해하기 쉽고 유용한 방법으로 (흔히 대용량) 데이터 집합을 분석”하는 방법으로 정의된다. 분류(예. 의사결정나무), 회귀분석, 군집화(예. k-군집화), 패턴 발견(예. 연관성규칙)과 같은 광범위한 기법들이 개발되어 있다.

운영관리 및 데이터 마이닝의 두 분야는 가치 있는 분석 기법들을 제공한다. 도전과제는 이 분야의 기법들을 프로세스 마이닝과 결합하는 것이다. 시뮬레이션을 예를 들어 보자. 과거 데이터에 근거하여 시뮬레이션 모델을 학습시키는 데 프로세스 마이닝 기법을 사용할 수 있다. 다음으로 운영적 지원을 제공하는 용도로 시뮬레이션 모델을 사용할 수 있다. 이벤트 로그와 모델은 서로 밀접하게 연계되어 있기 때문에, 과거를 재현하는 데 그 모델을 사용할 수도 있고, 진행되고 있는 데이터를 기반으로 현재 상태에서부터 미래를 보여주는 “고속재생(**fast forward button**)” 기능을 제공하여 시뮬레이션 을 시작할 수도 있다.

마찬가지로, 프로세스 마이닝을 “시각적 분석기법(**visual analytics**)”들과 결합하는 것도 바람직하다. 시각적 분석기법들은 크고 복잡한 데이터 집합을 잘 이해하기 위하여, 자동화된 분석을 상호작용할 수 있는 가시화 도구들과 결합하고 있다. 시각적 분석기법들은 비구조적 데이터의 패턴들을 살펴보기 위하여 인간의 놀라운 능력을 활용한다. 자동화된 프로세스 마이닝 기법을 상호작용할 수 있는 시각적 분석 기법과 결합하여, 이벤트 데이터로부터 더 많은 통찰력을 이끌어낼 수 있다.

C10: 비전문가들을 위한 사용성 개선

프로세스 마이닝의 목표 중 하나는 “생명력 있는 프로세스 모델(living process models)”을 생성하는 것이다. 즉, 어떤 아카이브에 완전히 종료된 정적인 모델보다는 하루 단위로 사용되는 프로세스 모델을 생성하는 것이다. 신규 이벤트 데이터는 새롭게 나타나는 행위를 발견하는 데 사용될 수 있다. 이벤트 데이터와 프로세스 모델을 연계함으로써 현재의 정적인 작업과 최신 작업이 업데이트된 모델에 반영할 수 있다. 그러므로 최종 사용자는 하루 단위로 생성된 프로세스 마이닝의 결과를 확인할 수 있다. 이러한 상호작용은 매우 가치 있는 일이며, 또한 직관적인 사용자 인터페이스를 필요로 한다. 이를 위해서는 복잡한 프로세스 마이닝 알고리즘을 숨긴 채, 자동으로 인자를 설정하고 적절한 분석 유형을 제안하는 사용자 친화적인 인터페이스가 필요하다.

C11: 비전문가들을 위한 이해도 개선

프로세스 마이닝 결과를 생성하는 것은 쉬운 일이지만, 그 결과가 실제로 유용하다는 것은 아니다. 사용자가 결과를 이해하는 데 문제가 있거나 또는 잘못된 결론에 도달할 우려가 있다. 이를 방지하기 위해서는 분석 결과를 적절한 표현을 사용하여 표현해야 한다(GP5 참고). 나아가, 결과의 신뢰도를 항상 명확하게 제시하여야 한다. 특정한 결론으로 판정하기에는 너무나 적은 데이터를 가지고 있을 수 있다. 실제로 기존의 프로세스 도출 기법들은 적합도가 낮거나 과적합이 되어도 이에 대한 경고를 하지 않는다. 어떤 결론을 내리기에는 분명히 너무 적은 데이터를 가지고 있을 때조차 항상 모델을 제시하고 있다.

맺음말

IEEE Task Force on Process Mining의 목표는 (1) 프로세스 마이닝 활용을 촉진하고, (2) 개발자, 컨설턴트, 비즈니스 관리자, 최종 사용자가 최신 기법들을 사용할 수 있도록 안내하고, (3) 프로세스 마이닝 연구를 촉

진하는 것이다. 이 매니페스토는 본 태스크 포스의 주요 원리와 의도를 기술한다. 본 프로세스 마이닝의 주제를 소개하고 적용 가이드라인(3장)과 도전과제(4장)를 제시하였다. 적용 가이드라인은 분명한 실수를 피하기 위하여 참고할 수 있으며, 도전과제는 연구와 개발에 직접 활용할 수 있다. 두 가지 모두 프로세스 마이닝의 성숙도 수준을 향상시키는 것을 목표로 한다.

마지막 부분의 용어집에서는 대표적인 용어를 설명하였다. 프로세스 마이닝 영역에서 워크플로우 마이닝, (비즈니스) 프로세스 마이닝, 자동화된 (비즈니스) 프로세스 도출, (비즈니스) 프로세스 인텔리전스와 같은 용어들이 중복되어 사용된다. 예를 들어, 가트너는 “자동화된 비즈니스 프로세스 도출”이라는 개념을 홍보하고 있고, Software AG는 통제 플랫폼을 가리키는 용어로 “프로세스 인텔리전스”를 사용한다. “워크플로우 마이닝”이라는 용어의 경우, 워크플로우 모델을 도출하는 것은 프로세스 마이닝에서 가능한 여러 응용 분야 중 하나일 뿐이므로 프로세스 마이닝을 대체하는 용어로 부적절하다. 마찬가지로, “비즈니스”라는 용어를 붙이는 것은 프로세스 마이닝의 특정 응용 분야로 범위를 좁히게 된다. 프로세스 마이닝의 매우 다양한 응용 분야가 있기 때문에(예를 들어, 하이테크 시스템의 이용을 분석하거나 웹사이트를 분석하는 등), “비즈니스”라는 용어를 붙이는 것은 부적절해 보인다. 프로세스 도출이 프로세스 마이닝 스펙트럼의 중요한 부분이지만, 많은 유스케이스 중의 하나일 뿐이다. 적합도 검사, 예측, 조직 마이닝, 사회적 네트워크 분석 등 프로세스 도출 이외에도 다양한 분야가 있다.

그림7은 위에서 언급된 일부 용어 간의 관계를 보여준다. 의사결정 지원에 필요한 실행 가능한 정보를 제공하는 모든 기술과 방법을 비즈니스 인텔리전스(Business Intelligence: BI)라고 부를 수 있다. (비즈니스) 프로세스 인텔리전스는 BI와 BPM이 결합된 개념으로 볼 수 있다. 즉, BI 기법들은 프로세스와 그 관리를 분석하고 개선하는데 마찬가지로 사용될 수 있다. 프로세스 마이닝은 시작 시에 이벤트 로그를 활용하는 프로세스 인텔리전스라고 구체화할 수 있다. (자동화된 비즈니스)

본 매니페스토는 “Business Process Management Workshops 2011 (Lecture Notes in Business Information Processing, Vol. 99, Springer-Verlag, 2011)”에 처음 발표되었고, 여러 언어로 번역이 되었다. 더 자세한 정보는 IEEE Task Force on Process Mining의 홈페이지(<http://www.win.tue.nl/ieeefpm/>) 참고.

프로세스 도출은 프로세스 마이닝의 기본적인 세 가지 유형 중 하나이다. 그림7은 대부분의 BI 도구들이 이 문서에 기술된 프로세스 마이닝 기능들을 제공하지 않는다고 간주하였기 때문에 일부 오해가 있을 수도 있다. BI라는 용어는 편의상 BI의 넓은 스펙트럼 중 일부분을 위한 특수한 도구와 방법을 지향한다고 간주하였다.

서로 다른 용어를 사용하는 상업적 이유가 있을 수 있다. 벤더들은 특정 측면(예. 도출 또는 지능화)을 강조하기를 원할 수도 있으나, 혼동을 피하기 위하여 이 매니페스토에서 언급된 분야에 대해서는 “프로세스 마이닝”을 사용하는 것이 더 바람직하다.

용어해설

- **작업(Activity):** 프로세스를 구성하는 단위 단계. 이벤트는 특정 프로세스 인스턴스에 대하여 한 작업의 시작, 종료, 취소 등을 나타낼 수 있다.
- **자동화된 비즈니스 프로세스 도출(Automated Business Process Discovery):** 프로세스 도출(Process Discovery) 참조.
- **비즈니스 인텔리전스(Business Intelligence(BI)):** 의사결정을 지원하기 위하여 데이터를 사용하는 도구와 방법들의 폭넓은 집합.
- **비즈니스 프로세스 인텔리전스(Business Process Intelligence):** 프로세스 인텔리전스(Process Intelligence) 참조.
- **비즈니스 프로세스 관리(Business Process Management(BPM)):** 지식을 정보기술과 결합하거나 지식을 경영과학과 결합하여 이를 운영적 비즈니스 프로세스에 적용하는 분야.
- **케이스(Case):** 프로세스 인스턴스(Process Instance) 참조.
- **개념 변화(Concept Drift):** 시간이 경과하면서 프로세스자 자주 변화하는 현상. 관찰

된 프로세스는 계절적 변화나 경쟁의 심화로 인하여 점진적으로 (또는 급진적으로) 변화하고, 그 결과 분석을 복잡하게 함.

- 적합도 검사(Conformance Checking): 로그에 기록된 대로 실체가 그 모델을 따르고 있는지, 그리고 그 반대로 따르고 있는지를 분석하는 작업. 적합도 검사의 목적은 불일치를 탐지하고 그 심각한 정도를 측정하는 것으로, 적합도 검사는 프로세스 마이닝의 세 가지 기본 유형 중의 하나임.

- 다조직 프로세스 마이닝(Cross-Organizational Process Mining): 서로 다른 조직에서 수집한 이벤트 로그들에 프로세스 마이닝을 적용하는 것.

- 데이터 마이닝(Data Mining): 예상치 못한 관계를 찾거나 데이터를 요약하기 위하여 (흔히 대용량) 데이터 집합을 분석하는 것.

- 이벤트(Event): 특정 프로세스 인스턴스에서 한 작업의 시작, 종료, 취소 등과 같이 로그에 기록되는 행위.

- 이벤트 로그(Event Log): 프로세스 마이닝을 위한 투입으로 사용되는 이벤트의 집합. 이벤트들이 하나의 분리된 로그 파일에 저장되어 있을 필요는 없음. (예를 들어, 이벤트가 서로 다른 데이터베이스 테이블들에 흩어져 있을 수도 있음.)

- 적합도(Fitness): 주어진 모델이 이벤트 로그에 나타난 행위들을 얼마나 잘 반영하는가를 측정하는 지표. 한 모델이 로그의 모든 흔적들을 시작부터 끝까지 모두 재현할 수 있다면, 그 모델은 완벽한 적합도를 가지는 모델임.

- 일반화(Generalization): 도출한 모델이 아직 나타나지 않은 행위들도 얼마나 잘 반영할 수 있는지를 측정하는 지표. "과적합" 모델은 충분히 일반화되지 않은 모델임.

- 모델 향상(Model Enhancement): 프로세스 마이닝의 세 가지 기본 유형 중 하나. 어떤 로그로부터 추출된 정보를 사용하여 프로세스 모델을 확장하고 개선함. 예를 들어, 타임스탬프를 확인한 후, 그 프로세스 모델에 이벤트 로그를 재현해 봄으로써 병목점을 식별할 수 있음.

- MXML: 이벤트 로그를 교환하기 위한 XML 기반의 포맷. 새로운 도구 독립적 프로세스 포맷으로서 XES이 MXML을 대체하고 있음.

- 운영적 지원(Operational Support): 진행 중인 프로세스 인스턴스를 모니터링하고 영향을 주기 위한 목적으로 이벤트 데이터를 온라인 분석하는 것. 세 가지 운영적 지원 활동으로 나눌 수 있음. 탐지(detect)는 관찰된 행위가 모델링된 행위로부터 벗어난 행위를 경우 알려주는 것이고, 예측(prediction)은 과거 행위에 근거하여 미래 행위를 예상하는 것(예. 남은 처리 시간을 예측)이며, 추천(recommend)은 특정 목표(예. 비용 최소화)를 실현하기 위하여 적절한 행위를 제안하는 것임.

- 정확도(Precision): 주어진 모델이 이벤트 로그에 나타난 행위로부터 너무 다르게 행동하지 않도록 하기 위하여 측정하는 지표. 정확도가 낮은 모델은 "과소적합"된 것임.

- 프로세스 도출(Process Discovery): 프로세스 마이닝의 세 가지 기본적인 유형 중 하나. 이벤트 로그에 근거하여 프로세스 모델을 학습하는 것. 예를 들어, α 알고리즘은 이벤트 집합 내의 프로세스 패턴을 확인함으로써 Petri net을 도출할 수 있음.

- 프로세스 인스턴스(Process Instance): 분석할 프로세스에 의해 처리된 개체들. 이벤트는 프로세스 인스턴스를 참조한다. 프로

세스 인스턴스의 예로는 고객 주문, 보험 청구, 대출 신청 등이 있음.

- 프로세스 인텔리전스(Process Intelligence): 비즈니스 프로세스 관리에 초점을 맞춘 비즈니스 인텔리전스의 한 분야.

- 프로세스 마이닝(Process Mining): 정보시스템에서 공통적으로 활용할 수 있는 이벤트 로그로부터 지식을 추출함으로써 실제 프로세스(예. 가정하지 않은 프로세스)를 도출, 모니터링, 개선하기 위한 기법, 도구, 방법들을 일컫는 용어.

- 표현적 편중(Representational Bias): 프로세스 마이닝 결과를 보여주거나 구축하기 위하여 선택된 대상 언어에 의한 발생하는 한계.

- 간결성(Simplicity): 오컴의 면도날(Occam's Razor)이 적용되는 척도. 예를 들어, 로그에 나타난 행위를 표현할 수 있는 모델 중에서는 가장 간결한 모델이 가장 훌륭한 모델이다. 간결성은 다양한 방법으로 정량화될 수 있음. (예. 모델에 포함된 노드나 아크의 개수).

- XES: 이벤트 로그를 위한 XML 기반의 표준. XML은 IEEE Task Force on Process Mining에 의하여 이벤트 로그를 위한 기본인 교환 포맷으로 채택됨 (cf. www.xes-standard.org).

본 매니페스토는 UNIST(울산과학기술대) 송민석 교수(msong@unist.ac.kr)와 경희대학교 정재운 교수(jjyung@khu.ac.kr)가 공동으로 번역하였습니다.

Authors

Wil van der Aalst
Arya Adriansyah
Ana Karla Alves de Medeiros
Franco Arcieri
Thomas Baier
Tobias Blickle
Jagadeesh Chandra Bose
Peter van den Brand
Ronald Brandtjen
Joos Buijs
Andrea Burattin
Josep Carmona
Malu Castellanos
Jan Claes
Jonathan Cook
Nicola Costantini
Francisco Curbera
Ernesto Damiani
Massimiliano de Leoni

Pavlos Delias
Boudewijn van Dongen
Marlon Dumas
Schahram Dustdar
Dirk Fahland
Diogo R. Ferreira
Walid Gaaloul
Frank van Geffen
Sukriti Goel
Christian Günther
Antonella Guzzo
Paul Harmon
Arthur ter Hofstede
John Hoogland
Jon Espen Ingvaldsen
Koki Kato
Rudolf Kuhn
Akhil Kumar
Marcello La Rosa
Fabrizio Maggi

Donato Malerba
Ronny Mans
Alberto Manuel
Martin McCreesh
Paola Mello
Jan Mendling
Marco Montali
Hamid Motahari
Nezhad
Michael zur Muehlen
Jorge Munoz-Gama
Luigi Pontieri
Joel Ribeiro
Anne Rozinat
Hugo Seguel Pérez
Ricardo Seguel Pérez
Marcos Sepúlveda
Jim Sinur
Pnina Soffer
Minseok Song
Alessandro Sperduti

Giovanni Stilo
Casper Stoel
Keith Swenson
Maurizio Talamo
Wei Tan
Chris Turner
Jan Vanthienen
George Varvaressos
Eric Verbeek
Marc Verdonk
Roberto Vigo
Jianmin Wang
Barbara Weber
Matthias Weidlich
Ton Weijters
Lijie Wen
Michael Westergaard
Moe Wynn