

# 조선 산업에서 프로세스 마이닝을 이용한 블록 이동 프로세스 분석 프레임워크 개발

이동하<sup>1</sup> · 배혜림<sup>2\*</sup>

<sup>1</sup>대우조선해양(주) 중앙연구소 / <sup>2</sup>부산대학교 산업공학과

## Analysis Framework using Process Mining for Block Movement Process in Shipyards

Dongha Lee<sup>1</sup> · Hyerim Bae<sup>2</sup>

<sup>1</sup>Central R&D Institute, Daewoo Shipbuilding and Marine Engineering Co., Ltd.

<sup>2</sup>Department of Industrial Engineering, Pusan National University

In a shipyard, it is hard to predict block movement due to the uncertainty caused during the long period of shipbuilding operations. For this reason, block movement is rarely scheduled, while main operations such as assembly, outfitting and painting are scheduled properly. Nonetheless, the high operating costs of block movement compel task managers to attempt its management. To resolve this dilemma, this paper proposes a new block movement analysis framework consisting of the following operations: understanding the entire process, log clustering to obtain manageable processes, discovering the process model and detecting exceptional processes. The proposed framework applies fuzzy mining and trace clustering among the process mining technologies to find main process and define process models easily. We also propose additional methodologies including adjustment of the semantic expression level for process instances to obtain an interpretable process model, definition of each cluster's process model, detection of exceptional processes, and others. The effectiveness of the proposed framework was verified in a case study using real-world event logs generated from the Block Process Monitoring System (BPMS).

**Keywords:** Process Mining, Fuzzy Mining, Trace Clustering, Shipyard, Shipbuilding, Block Movement, Transportation, Event Logs

### 1. 서론

조선 산업의 제조 공정은 기본적으로 철판을 절단/벤딩하는 가공작업을 시작으로 블록을 조립하는 선각 작업, 의장, 도장 작업 그리고 탑재 작업 순으로 이루어진다. 선박 설계 후, 생산을 위해서는 선박을 작업장 내에서 조립할 수 있는 적절한 크기의 블록들로 나누는데, 이를 블록 분할(block division) 작업이라고 한다. 또한, 선박은 나누어진 블록을 기준으로 건조되

는데, 이를 블록 건조 공법이라고 한다. 조립 작업장 내에서 용접 작업에 의해 조립된 조립 블록(assembly block)들은 선형의 장, 도장 등의 작업을 거쳐 도크 주변의 선형탑재 지역으로 옮겨진다. 도크에서 블록들이 선박으로 조립되는 시간을 줄이기 위해서 작업장 내에서 제작된 조립 블록들은 선형탑재 지역에서 대형 블록으로 조립되는데 이렇게 만들어진 블록을 선형탑재 블록(pre-erection block)이라고 부른다. 최종 선박은 도크에서 블록과 선형탑재 블록들의 탑재에 의해서 만들어진다. 도

이 논문은 2012년도 교육과학기술부의 재원으로 한국연구재단의 기초연구 사업(No. 2012R1A1A2008335)과 (주)대우조선해양의 지원을 받아 수행된 것입니다.

\* 연락처 : 배혜림, 609-736 부산광역시 금정구 장전동 부산대학교 산업공학과, Fax : 051-512-7603, E-mail : hrbae@pusan.ac.kr  
2013년 2월 4일 접수; 2013년 4월 25일 수정본 접수; 2013년 5월 14일 게재 확정.

크에서 최종 배의 모습을 갖춘 선박은 안벽으로 옮겨져 마무리 의장, 도장 작업을 하게 된다. 본 논문은 이러한 선박의 생산 공정 중, 여러 조립 작업장에서 제작된 조립 블록들이 선행 탑재 작업장 혹은 도크에서 탑재될 때까지의 이동 데이터에 관심을 둔다.

현재 우리나라 조선소에서 생산계획을 수립할 때, 주요 공정인 조립, 의장, 도장 작업은 일정계획을 수립하지만, 블록 이동은 모든 경우에 대한 예측이 힘들어서 일정계획을 수립하지 않는 것이 일반적이다. 일정계획을 수립하는 대신, 주요 작업 공정의 생산 계획에 따른 블록 이동 빈도를 예측하고 대비하는 수준에 머무르고 있다. 그러나 실제 생산에서는 예측된 것보다 훨씬 많은 블록의 이동이 존재한다. 조선 산업의 특성상 생산 계획 단계에서 관리하는 작업의 수준보다 세분화된 작업들을 기준으로 생산이 이루어질 뿐만 아니라, 주요 작업 이외에도 다양한 이유로 블록 이동이 발생하기 때문이다. 예를 들어, 특정 작업을 마치고 다음 공정의 작업장으로 이동하기 전 적치장에서의 이동이 발생하고, 이 경우도 적치장의 상황에 따라 다음 공정의 작업에 유리한 적치장으로의 이동이 가능한 경우와 그렇지 못한 경우가 있다. 후자의 경우는 이후 적치장 간 이동도 발생하게 된다. 적치장 내에서도 다른 블록의 이동 및 배치를 위해서 추가 이동이 발생하기도 한다. 또한 특정 공정의 작업이 완료되지 않았는데도 다른 블록의 긴급 작업 대응을 위해서 이동이 발생하기도 하고, 검사 및 재작업을 위해서 추가 이동이 발생하기도 한다. 사외 아웃소싱(outsourcing)을 통해 제작된 블록은 야드(yard)내 입고 지역으로 들어오고, 사외에서 작업한 이후 공정을 시작으로 작업 및 블록 이동이 발생하게 된다. 조립블록의 경우만 하더라도 무게가 수백 톤(ton)에 이르고, 야드의 면적이 넓은 관계로 트랜스포터(transporter)를 이용한 수송비용이 많이 소모된다. 따라서 작업 관리자들은 불필요하고, 불합리한 블록 이동을 막기 위해 블록 이동 프로세스를 분석하고 관리해야만 한다.

본 논문에서는 이러한 조선 산업의 블록 이동 프로세스를 프로세스 마이닝(process mining) 기법을 이용하여 분석하는 프레임워크(framework)를 제안한다. 블록 이동에 대한 데이터는 블록 공정 모니터링 시스템(BPMS : Block Process Monitoring System)의 로그 데이터를 활용한다. BPMS 데이터는 블록이동에 대한 실적 정보를 가지고 있다. 생산 부서의 작업자가 블록 이동을 요청 하게 되고, 실제 블록 이동이 발생할 때마다 트랜스포터 조정자 혹은 블록 운반 관리자가 처리 결과를 시스템에 입력하게 된다. BPMS는 블록의 상하 혹은 하차 정보만 입력하면 다른 부가 정보는 자동으로 입력될 정도로 사용자 관점의 높은 편의성을 제공하고, 데이터의 신뢰도가 높다. 이 데이터를 활용하면 생산 계획 데이터 기준의 예측이 아니라, 실적 데이터 기준의 실제 생산에서 발생한 블록 이동 패턴을 파악하고 분석할 수 있다. 그리고 본 논문에서 제안한 프레임워크는 실제 생산의 블록 이동에 대한 분석을 용이하게 하고, 해당 결과를 통해 개선, 관리해야 하는 부분을 확인할 수 있는 장점을 가진다.

## 2. 관련 연구

프로세스 마이닝은 기업정보시스템에서 프로세스 실행 과정에서 발생한 이벤트 로그로부터 의미 있는 지식을 추출해 내는 과정이다. 즉, 프로세스의 액티비티 실행 과정에서 누적된 기록을 통하여 프로세스의 개선이나 설계에 필요한 유용한 지식을 추출하는 것이다. 프로세스 저장소의 실행 결과 및 이벤트에 관한 로그를 추출하여 기존의 통계 기법, 인공지능 기법, 사회적 네트워크 기법 등을 이용하여 분석함으로써(Medeiros *et al.*, 2007; Jansen-Vullers *et al.*, 2006; Rozinat and Aalst, 2006; Aalst *et al.*, 2005), 프로세스 모델을 유도하거나, 업무의 상관관계, 작업자의 업무 전달 관계 등 프로세스 수행 상의 특징을 분석한다(Aalst and Basten, 2002).

프로세스 마이닝은 데이터 마이닝(Data Mining)과 비즈니스 프로세스 관리(BPM : Business Process Management)와도 비교된다(Weerdt, 2012). 데이터 마이닝도 대용량의 복잡한 데이터로부터 유용한 정보를 추출하고 의미 있는 규칙이나 패턴을 찾는 과정으로 정의 되므로(Berry and Linoff, 1997), 프로세스 마이닝과 같이 데이터로부터 정보를 추출해 낸다는 공통점이 있다. 그러나 프로세스 마이닝은 로그 데이터의 시간 정보를 활용해서 프로세스를 발견하고, 해당 프로세스와 관련된 분석에 초점을 맞춘다는 특징을 가진다. 따라서 현재 개발된 프로세스 마이닝 기법을 이용하면 데이터의 많은 전처리 없이 쉽게 프로세스 모델을 얻을 수 있고, 해당 프로세스에 대한 분석이 가능한 장점이 있다. BPM과 비교하면 비즈니스 프로세스에 대한 통찰력을 얻는다는 공통점은 있지만, 프로세스 마이닝은 로그 데이터로부터 다양한 기법을 이용하여 프로세스 모델을 추출하는 것에서부터 시작하기 때문에 기존에 정의된 프로세스 모델을 활용하는 BPM과는 출발점이 다르다. 프로세스 모델링은 BPM에서도 시작 단계로, 중요하게 인식 되어 프로세스간의 유사성을 측정하거나(Yahya *et al.*, 2011) 프로세스의 구조 제약조건과 사회적 관계를 이용하는(Yu *et al.*, 2012) 방안 등이 제시되었다.

프로세스 마이닝에 대한 연구는 대부분 프로세스 로그로부터 프로세스 모델을 추출하는 기법에 관한 것이었다. 아인트호벤 공대에서 개발한 오픈 프레임워크인 ProM은 플러그인(plug-in) 형태로 구현되어 있다(Aalst *et al.*, 2009). 이 프레임워크는 ERP나 BPM 시스템의 프로세스 로그를 임포트(import)하여 프로세스 모델 추출, 프로세스 모델 분류, 프로세스 패턴 추출, 작업자 간 업무 전달 형태 등을 분석할 수 있는 환경을 제공한다. 본 연구에서도 프로세스 마이닝 기법의 활용을 위해 ProM을 이용하였다.

조선 산업에서도 최근 데이터 마이닝이나 프로세스 마이닝을 이용한 생산 정보에 대한 분석에 관심을 가지기 시작 하였다. 조선해양 분야의 세계적인 선두 기업인 국내 업체들은 최근 몇 년 전까지는 ERP 시스템 구축과 안정화에 집중 하였고, 생산 정보의 모니터링 관점에서 최신 IT기술을 적용하거나, 정

보시시스템을 구축하는 활동을 주로 하였다. 이러한 정보시스템의 안정화로 데이터가 축적되자, 최근에는 데이터 분석에도 관심을 가지게 되었다.

본 연구에서 대상 문제로 정한 블록의 이동과 배치 문제는 대형 구조물을 만드는 조선 산업의 특성상 늘 중요한 문제로 다루어져 왔다. 그러나 상위 생산계획 단계의 관리 수준에서 작업을 정의하고, 특정 적치장 혹은 작업장의 계획 관점에서 최적화에 대한 연구가 주로 수행되었다(Lee *et al.*, 2011; Park and Seo, 2012). 실제 생산에서는 보다 세분화된 수준의 공정 단위로 작업이 이루어지고, 다양한 예외 상황에 의해서 블록 작업과 이동이 발생하므로 실적 데이터 기준의 분석과 연구가 더 많이 요구된다.

최근 블록 이동에 대한 실적 정보를 가지고 해당 블록이 지나간 작업장 데이터를 기준으로 데이터 마이닝 기법을 적용해서 블록을 군집화 하고 프로세스를 분석한 연구가 있다(Lee *et al.*, 2013). 본 논문에서는 데이터마이닝 기법을 넘어서, 다양한 프로세스 마이닝 기법을 이용해서, 전체 프로세스를 이해하고, 관리할 수 있는 수준의 프로세스 모델을 찾고, 예외 활동을 감지하고 분석하는 전체적인 방법론을 제안한다. 프로세스 마이닝 연구 관점에서 프로세스를 도출하거나, 특정 분석을 위한 알고리즘을 개발하는 것이 아니라, 실제 제조업의 데이터를 활용하여 프로세스 분석 프레임워크를 제안한다는 데 의미가 있다.

### 3. 제안 프레임워크

본 논문에서는 프로세스 마이닝을 이용한 블록 이동 프로세스 분석 프레임워크를 제안한다. 블록 공정 모니터링 시스템의 실적 데이터를 이용해서 실제 생산단계에서 발생한 블록의 이동 패턴을 찾고, 프로세스 모델을 정의하고, 예외 공정에 대한 분석과 관리를 지원하는 것이 목적이다. 블록 공정 모니터링 시스템의 이벤트 로그 데이터 중, 본 연구에서 대상으로 하는 주요 데이터는 <Table 1>과 같다. <Table 1>에서는 단지 한 개의 블록에 대한 로그 정보를 추출하여 예로 제시하였지만, 데이터베이스에는 동시에 작업 되는 많은 블록에 대해서 이벤트

가 발생하는 순서에 따라 로그 정보가 축적되고, 블록의 이동 거리, 시간, 위치 등에 대한 정보도 저장된다. 본 논문에서는 문제의 특성상 블록의 이동 패턴을 찾고, 프로세스 모델을 정의하고 관리하는 방법에 더 많은 관심을 가진다.

제안한 프레임워크는 <Figure 1>과 같이 데이터 전처리 단계인 단계 0(Step 0)을 제외하고 목적에 따라 네 단계(Step)로 구성된다. 전처리 단계에서는 이동 작업장에 대한 데이터가 누락된 경우, 동일 시각에 데이터가 중복 입력된 경우 등에 대한 데이터 수정 혹은 삭제 작업이 필요하다. 또한 분석에 포함하고자 하는 대상 필드를 정의하고, 데이터베이스에 코드로 입력된 데이터를 분석이 용이한 표현으로 변환하는 작업 등이 요구된다.

단계 1(Step 1)은 분석하고자 하는 데이터로부터 전체 프로세스를 이해하는 단계이다. 일반적으로 정보시스템의 로그 데이터를 이용해서 프로세스 마이닝을 하게 되는 경우는 프로세스가 매우 복잡하고, 원천 데이터를 통한 해석이 직관적이지 않은 경우이다. 따라서 분석하고자 하는 수준으로 데이터의 의미론적 표현을 변경한 후, 전체 데이터 속에서 중요한 프로세스와 상대적으로 덜 중요한 프로세스를 구분하여 전체를 조망하는 작업이 가장 먼저 선행되어야 한다. 퍼지 마이닝(Fuzzy Mining) 기법은 각 노드에 해당하는 활동(activity)과 활동 간의 관계(relation) 정보를 이용하여 프로세스 내 이들의 역할에 따라 중요한 활동과 상대적으로 덜 중요한 활동을 구분하고, 덜 중요한 활동들은 추상화된 군집(abstracted clusters)을 형성해서 전체 프로세스를 쉽게 확인할 수 있게 한다(Günther and Aalst, 2007). 그러므로 단계 1에서는 분석하고자 하는 전체 데이터에 대해서 퍼지 마이닝과 데이터의 의미론적 표현(semantic expression) 수준을 조정하는 반복 활동을 통해 전체 프로세스의 개요를 이해할 수 있는 수준으로 만드는 작업이 요구된다.

단계 2(Step 2)는 관리할 수 있는 프로세스를 얻기 위해서 로 그들을 군집화 하는 단계이다. 전체 분석 데이터를 통해서 하나의 프로세스 모델을 만들고 이해하는 것이 단계 1이라면, 로 그들을 군집화 한 후 각각의 프로세스 모델을 만드는 것이 단계 2의 작업이라고 하겠다. 일반적으로 프로세스 관리자나 작업 관리자는 데이터 전체를 대상으로 하나의 프로세스 모델을 보유하기 보다는 관리할 수 있는 수준의 여러 모델을 가지는

Table 1. Example of Event Logs for a Block

Project	Block	Workshop	Moving Datetime	Organization	Transporter
A	Block 1	Assembly Shop	Oct-26 09:31	BB	Z
A	Block 1	Outfitting Shop	Nov-24 22:20	BB	Y
A	Block 1	Quay+Stock Area	Nov-26 21:20	CC	X
A	Block 1	Blasting Shop	Dec-08 06:10	DD	Y
A	Block 1	Painting Shop	Dec-09 11:10	DD	Y
A	Block 1	Painting Area (Outdoor)	Dec-13 22:35	DD	Y
A	Block 1	Stock Area	Dec-20 20:14	DD	X
A	Block 1	HEAVY+PE Area	Dec-21 08:46	DD	X

것을 선호한다. 로그 데이터의 군집별 특징을 찾고, 해당 특징 별로 프로세스 모델을 정의하면 프로세스 관리를 보다 효율적으로 할 수 있기 때문이다. 자취 군집화(Trace Clustering) 기법은 로그들을 군집화 하는 다양한 방법을 제공한다(Song *et al.*, 2009). 본 프레임워크에서는 자취 군집화 후, 군집 결과를 각 군집별 시퀀스 다이어그램(Sequence Diagram)으로 확인하고 군집별 특징을 비교할 것을 제안한다. 또한 관리할 수 있는 수준의 특징이 뚜렷한 군집화 결과를 얻을 때까지 다양한 군집화 알고리즘을 적용해 보아야 한다.

수 있다. 이러한 표준 프로세스 모델과 예외 프로세스의 정의는 실적 데이터에 대한 결과 이해와 현재 프로세스 진행에 대한 모니터링에 중요한 자료로 활용될 수 있다.

### 4. 사례 연구

블록 공정 모니터링 시스템의 실적 데이터 중 1개의 프로젝트 데이터를 이용하여 조립 블록 이동 프로세스를 분석 하였다. 해당 프로젝트는 중복 데이터를 제외하고 불완전한 데이터를 제거하는 등의 데이터 전처리 후에 총 3167개의 이벤트(event) 로그들이 존재 했고, 212개의 인스턴스(instance)들로 구성 되었다. 블록 이동 프로세스 분석 문제에서 인스턴스는 조립 블록이 된다. 하나의 조립 블록이 선행단계 작업장 혹은 도크에서 탑재 될 때까지의 이동 프로세스가 하나의 인스턴스가 되는 것이다.

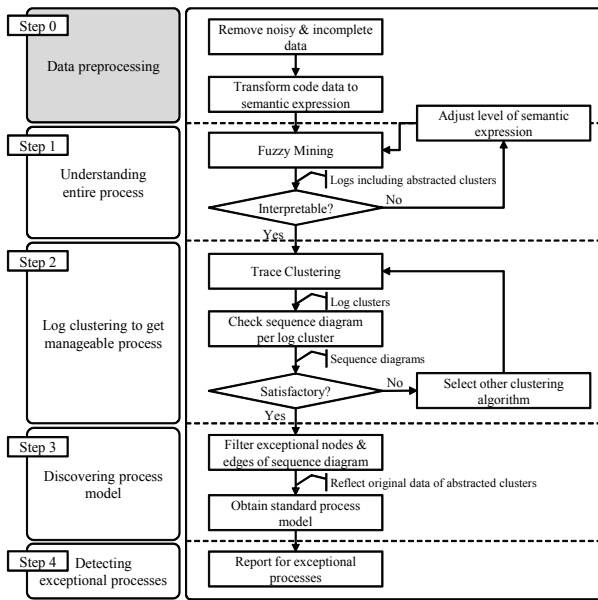


Figure 1. Framework for New Approach

단계 3(Step 3)은 각 로그 군집으로부터 프로세스 모델을 발견하는 단계이다. 단계 2에서 얻은 각 군집별로 프로세스 모델을 정의하는 것이다. 이 단계에서는 단계 1에서 퍼지 마이닝 결과로 얻은 덜 중요한 활동들에 대한 추상화된 군집 데이터를 사용하지 않고, 원래 데이터의 표현으로 재 반영한 후 모델을 정의할 것을 제안한다. 단계 2까지는 덜 중요한 활동들에 대해서 추상화를 해 두면 전체 프로세스를 이해하고 로그 군집화를 하는데 도움이 되지만, 단계 3에서 정의한 모델은 실제 프로세스 관리에 활용하기 위한 것이므로 원래 데이터의 표현 수준으로 바꾸는 것이 필요하다. 각 로그 군집에서 예외적인 활동에 대한 필터링 작업을 통해 표준 프로세스를 얻을 수 있다. 이는 대상 문제에 따라서 차이가 있으므로 사례연구에서 블록 이동 프로세스 분석의 경우 어떤 기준을 적용했는지 설명하도록 하겠다.

단계 4(Step 4)는 예외 프로세스를 감지하고 정리하는 단계이다. 단계 3에서 얻은 표준 프로세스 모델과 비교해서 예외 프로세스가 얼마나 발생했는지, 어떤 유형들이 존재하는지 레포팅(reporting)한다. 이를 기반으로 개선해야 하는 부분을 찾을 수 있고, 관심 있게 관리해야 하는 부분에 대한 정보를 얻을

#### 4.1 전체 프로세스 이해(Step 1. Understanding entire process)

전체 데이터로부터 프로세스를 이해하기 위해서 퍼지 마이닝의 수행과 데이터의 의미론적 표현 수준 조정 작업을 반복할 것을 제안 하였다. 본 사례 연구에서도 원천 데이터에 대해서 퍼지 마이닝만으로는 전체 프로세스를 이해하기에는 어려움이 있었다. 따라서 데이터의 의미론적 표현 수준을 조정하는 작업을 병행 하였고, 이는 퍼지 마이닝과 함께 간단하고 이해할 수 있는 전체 프로세스 모델을 얻는 것을 가능하게 하였다. 예를 들어 각 조립 작업장 코드 혹은 명칭을 단순히 조립 작업을 의미하는 ‘Assembly Shop’으로 표현함으로써 의미론적 표현 수준을 단순화 할 수 있다. 이는 각 조립 작업장에서 조립된 블록들이 어떻게 이동하는지에 관심이 있는 것이 아니라, 조립 공정 후에 어떤 이동들이 있는 것인지를 분석하고자 한다면 모델의 노드 수를 줄일 수 있으므로 간단한 모델을 얻는데 도움이 된다.

데이터의 의미론적 표현 수준 조정 작업과 퍼지 마이닝의 반복 작업 후에 최종 얻은 모델을 분석한 내용이 <Figure 2>와 같다. <Figure 2>에서 사각형 노드(node)는 설정한 임계값보다 활동 비율이 높은 주요 활동을 그대로 보여주고, 팔각형 노드는 설정한 임계값보다 활동비율이 낮은 활동들을 추상화한 군집을 표현한다(Günther and Aalst, 2007). 추상화된 각 군집은 <Figure 3(a), 3(b)>와 같이 상세화된 그래프를 통해 별도로 확인할 수 있다. <Figure 3(a), 3(b)>에서 음영이 있는 모서리가 둥근 사각형 노드가 추상화된 군집에 포함된 활동이고, 이들 활동과 연결된 <Figure 2>의 노드들이 투명한 직사각형과 팔각형으로 표현된다. <Figure 2>에서 볼 수 있는 것처럼, 주요 프로세스의 흐름을 찾고 추상화된 각 군집에 대해서 세부 활동의 확인을 통해 군집별 이름을 부여 하였다. 주요 프로세스(Main Process)는 조립 검사 혹은 조립 후 적치장으로의 이동이 군집화 되어 추상화

된 ‘Cluster 42’를 시작으로 ‘Outfitting Shop’, ‘Painting Shop’, ‘Painting Area(Outdoor)’, ‘HEAVY+PE Area’로의 이동으로 요약되었다. ‘Cluster 42’ 전에 ‘Stock Area’를 거치는 경우도 있는데, 프로세스 흐름의 빈도가 ‘Cluster 42’에서 시작하는 경우가 훨씬 많음을 ‘Start’ 노드와 ‘Cluster 42’ 사이의 연결선(link) 굵기로 확인할 수 있다. 주요 프로세스를 기준으로 왼쪽에 위치한 ‘Cluster 36’, ‘Cluster 30’, ‘Cluster 32’는 적치장으로 사용된 경우가 각 특성에 맞게 군집화 되어 추상화된 것이다. ‘Cluster

36’은 선형 탑재 지역이 적치장으로 사용된 경우, ‘Cluster 30’은 주요 공정 관련 전/후 특정 적치장으로 이동된 경우, ‘Cluster 32’는 안벽 주변 지역이 적치장으로 사용된 경우에 해당하는 것으로 파악되었다. 주요 프로세스를 기준으로 오른쪽에 위치한 ‘Cluster 35’와 ‘Cluster 40’은 특정 공정을 포함한 활동이 추상화된 것이다. ‘Cluster 35’는 옥외 작업장에서 의장을 하는 경우와 관련된 작업들로, ‘Cluster 40’은 도장 작업 전에 전처리 (Blasting) 작업을 포함한 관련 이동들로 구성되었다.

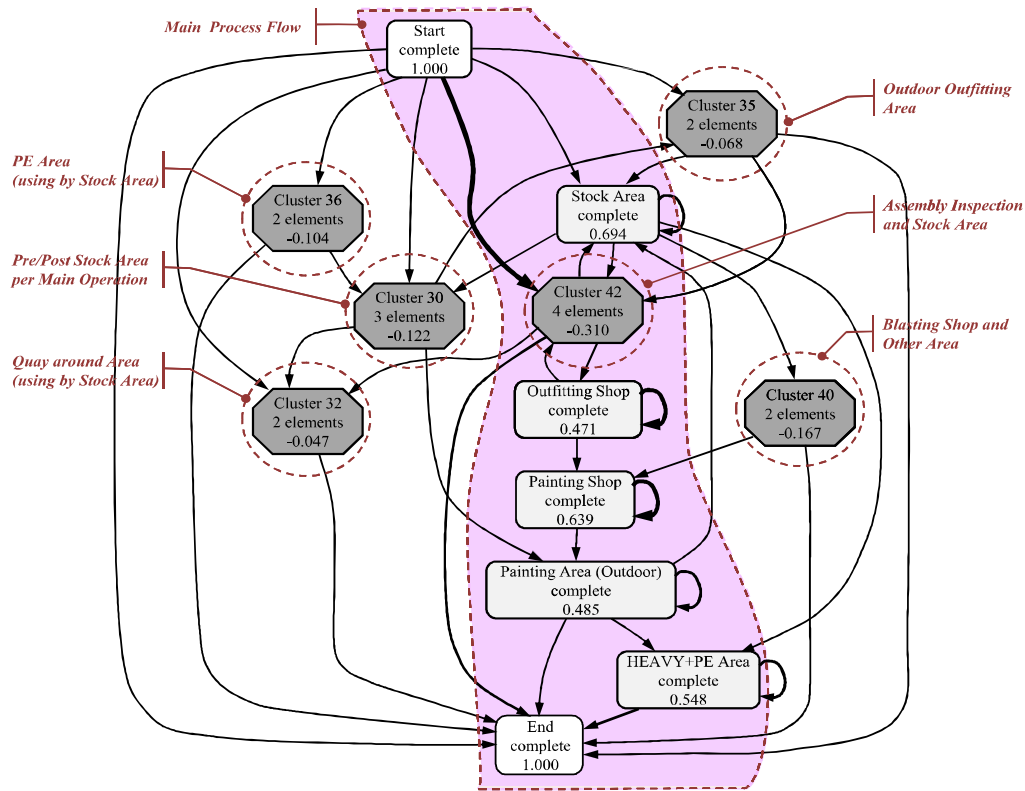


Figure 2. Understanding Fuzzy Mining Output

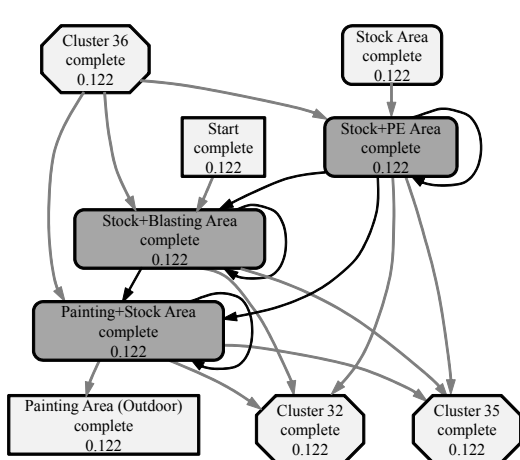


Figure 3(a). Detailed Graph for ‘Cluster 30’

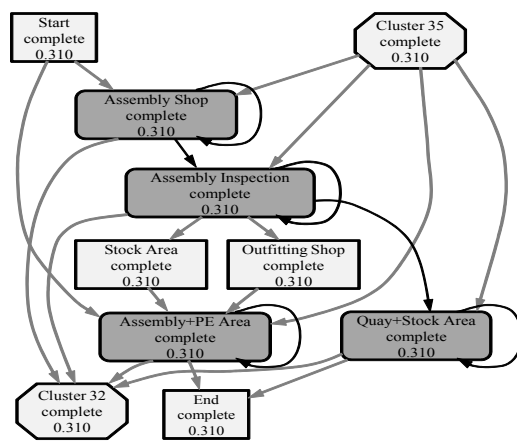


Figure 3(b). Detailed Graph for ‘Cluster 42’

결과적으로 데이터의 의미론적 표현 수준 조정 작업이 전체 주요 활동의 분석 수준을 결정하고, 퍼지 마이닝의 추상화 작업이 해당 활동에 대한 이해를 중요 수준별로 이해할 수 있게 한다고 하겠다. 본 논문에서는 퍼지 마이닝을 통해 추상화된 군집들을, 예를 들어 ‘Cluster 30’과 ‘Cluster 42’ 등을, 이후 단계 2의 입력 값으로 그대로 사용하였다. 상대적으로 덜 중요한 데이터의 경우는 추상화된 군집의 형태로 두고 분석을 하는 것이 전체적인 빠른 이해를 돕기 때문이다. 따라서 단계 1에서 제안한 방법은 정보시스템의 로그 데이터로부터 관심 있는 프로세스와 활동을 찾아내고, 전체 프로세스의 개요를 이해할 수 있는 좋은 출발점이 된다.

4.2 로그 군집화(Step 2. Log clustering)

단계 1에서 전체 분석 데이터를 통해 중요 프로세스의 개요를 확인한 후에 관리할 수 있는 수준의 프로세스를 얻기 위해

서 로그 군집화를 수행하였다. 자취 군집화는 전체 이벤트 로그들을 동질의 하위 집합(subset)들로 나누고 각 하위 집합에 대해서 프로세스 모델을 획득하는 것을 의미한다.

본 대상 문제에서는 군집화 알고리즘(algorithm)들 중에서 병합 계층 군집화(agglomerative hierarchical clustering)와 k-평균 군집화(k-means clustering)를 사용했다. 먼저 전체 데이터 로그에 대해서 의미 있는 하위 로그 군집의 개수를 찾기 위해서 병합 계층 군집화를 수행하였다. 그리고 나서, 병합 계층 군집화를 통해 확인한 의미 있는 군집 수를 k 값으로 설정하고, k-평균 군집화를 순차적으로 수행 하는 것이 효과적이었다. 전체 로그의 최종 군집화 결과는 각 군집별 시퀀스 다이어그램으로 확인 하였다.

사례 연구에서는 병합 계층 군집화를 통해 3개의 군집으로 나눌 것을 결정하였고, k-평균 군집화를 통해 3개의 군집을 구성하였다. 군집화를 통해 얻은 군집 0(Cluster 0), 군집 1(Cluster 1), 군집 2(Cluster 2)는 각각 인스턴스를 84개, 71개, 57개를 가

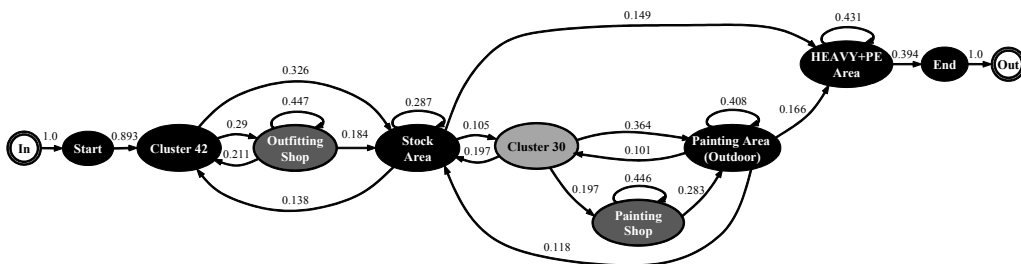


Figure 4(a). Sequence Diagram of ‘Cluster 0’ after Log Clustering

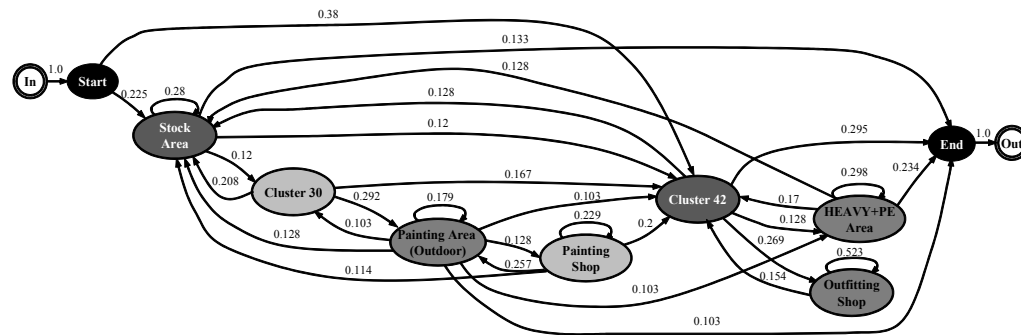


Figure 4(b). Sequence Diagram of ‘Cluster 1’ after Log Clustering

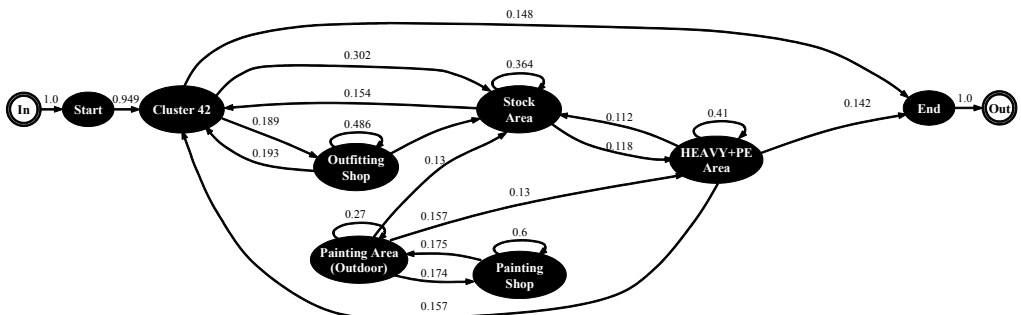


Figure 4(c). Sequence Diagram of ‘Cluster 2’ after Log Clustering

지게 되었고, 군집별 이벤트 수는 각각 1310개, 556개, 1301개로 구성되었다.

각 로그 군집별 프로세스에 대한 시퀀스 다이어그램은 <Figure 4(a), 4(b), 4(c)>와 같다. 시퀀스 다이어그램 내의 노드 색상은 해당상태로 변이되는 확률이 높은 정도에 따라 짙게 표현 되고, 연결선 위의 값은 마르코프 체인(Markov Chain)의 확률 값을 나타낸다(Veiga and Ferreira, 2010). 3개의 시퀀스 다이어그램은 모두 노드와 연결(link)의 하위 임계치가 0.1로 설정되어 정제(filtering)되었다. 군집 0의 <Figure 4(a)>는 단계 1에서 확인한 주요 프로세스를 가장 잘 따르는 흐름을 보인다. 조립작업 후 관련 활동들로 구성된 'Cluster 42'를 시작으로 'Outfitting Shop'을 거쳐 'Stock Area'에 이른다. 도장 작업 전에 관련 작업 주요 적치장으로서의 이동을 포함하는 'Cluster 30'으로서의 이동이 있고, 이후 도장 작업장인 'Painting Shop' 혹은 옥외 도장 지역인 'Painting Area(Outdoor)'를 거쳐서 'HEAVY+PE Area' 지역에서 최종 탑재가 된다. 군집 2의 <Figure 4(b)>는 적치장 지역인 'Stock Area'에서 시작하는 흐름과 'Cluster 42'에서 시작하는 흐름으로 구분된다. 'Stock Area'에서 시작하는 경우는 사외에서 블록이 제작되어 사내로 입고된 후 적치장에 위치하다가 도장 작업을 거쳐 탑재 되는 경우로 이해된다. 'Cluster 42'에서 시작하는 경우는 군집 1과의 차이가 가장 작업 후에 도장 관련 작업 없이 바로 탑재되는 특징을 가진다. <Figure 4(c)>의 군집 3도 군집 1과 같이 'Cluster 42'에서 시작하지만, 의장 관련 작업과 도장 관련 작업이 순차적으로 일어나지 않고, 필요 작업이 적치장을 거쳐 각각 일어나고 탑재되는 것을 확인할 수 있다. 이렇게 각 로그 군집은 프로세스 관점에서 뚜렷한 차이를 보였다.

본 사례 연구를 통해 확인 할 수 있는 것처럼 로그 군집화를 한 후 프로세스 모델을 획득하면 각 로그의 군집별 특성에 따른 프로세스 모델을 확인 할 수 있다. 이것은 블록 이동에 대한 패턴을 찾고 관리할 수 있는 프로세스 모델을 획득하는 것에 도움을 준다.

**4.3 프로세스 모델 발견(Step 3. Discovering process model)**

단계 3에서는 단계 2에서 획득한 로그 군집별로 표준 프로세스 모델을 발견하고 정의하는 과정이다. 단계 2에서 수행한 로그 군집화 결과는 이미 군집별 시퀀스 다이어그램을 통해 그 특징을 확인 하였다. 그러나 단계 2의 시퀀스 다이어그램은 실제 데이터를 기반으로 프로세스 모델이 만들어졌기 때문에 많은 예외 상황까지도 그대로 포함하고 있다. 따라서 프로세스 관리 차원의 표준 모델로 활용하기에는 어렵다. 그러므로 본 문제의 특성을 고려하여 예외 상황들을 정제(filtering)할 수 있는 기준을 마련하고, 군집별 표준 프로세스 모델을 정의하였다. 이 단계에서는 단계 1에서 퍼지 마이닝에 의해 추상화된 'Cluster' 표현을 사용하지 않고 초기 데이터 수준으로 의미론적 표현을 재 반영하여 모델을 구성하였다. 예를 들어, Figure 3(b)에서 제시된 예에서 'Cluster 42'로 추상화 되었던 활동들

을 다시 'Assembly Shop', 'Assembly Inspection', 'Assembly+PE Area', 'Quay+Stock Area' 활동으로 각각 되돌려 사용하는 것이다. 이것은 실제 데이터 관리 수준에서 프로세스 모델을 정의하고 관리하기 위한 것이다. 이전까지는 프로세스를 이해하고 로그 군집별 특성을 파악하기 위한 방법들이라면, 이 단계부터는 실제적인 활용을 위한 표준 프로세스 모델을 획득하는 방법을 제안하는 것이다.

본 사례 연구에서는 불필요한 이동 및 불합리한 이동에 해당되는 예외 활동을 3가지 경우(Case)로 정의 하였다. Case 1은 '재 취급(Re-handling Movement)', Case 2는 '역 물류(Reverse Movement)', Case 3은 '불필요 이동 (Unnecessary Movement)'이다. '재 취급'은 데이터 상으로 동일 지역 내에서의 이동으로 정의 하였고, '역 물류'는 지나온 주요 공정 작업장으로 되돌아가는 경우로 간주 하였다. 주요 공정의 작업장이 아닌 특정 적치장으로 되돌아가는 경우는 Case 2에 포함시키지 않았다. '불필요 이동'은 적치장에서 다른 적치장으로 이동하는 경우에 해당된다. 단, 주요 공정 전과 후에 특정 적치장으로 이동하는 경우는 해당 작업의 빠른 흐름을 위한 것이므로 적치장 간 이동이라도 Case 3의 경우에서 제외했다. 상기 정의한 데이터들은 실제 생산 상황별로 여러 가지 이유로 발생했을 지라도, 작업 관리자 입장에서는 향후 발생하지 않기를 원하는 예외 상황 들이다. 따라서 표준 모델에서는 예외상황들을 제외하고, 별도로 관리할 수 있는 방안이 필요하다.

상기 정의한 Case 1, 2, 3의 예외 활동들을 정제한 후, 군집별 프로세스를 시퀀스 다이어그램으로 표현한 결과가 <Figure 5(a), 5(b), 5(c)>와 같다. 이들 시퀀스 다이어그램은 모두 노드와 연결(link)의 하위 임계치가 0.15로 설정되어 정제 되었다. <Figure 5(a), 5(b), 5(c)>에서 확인할 수 있는 것처럼, 퍼지 마이닝에 의해 추상화된 'Cluster 42', 'Cluster 30' 등의 노드 표현이 사라지고, 현재 데이터 수준으로 더욱 직관적인 프로세스 해석이 가능해 졌다. 역시 <Figure 5(a)>의 군집 0의 표준 프로세스 모델이 단계 1에서 파악한 주요 프로세스 흐름을 가장 잘 따르고 있다. 조립 작업장인 'Assembly Shop'에서 조립이 완료 되고 나서, 의장 작업을 위해 'Outfitting Shop'으로 이동하고, 이후 'Blasting', 'Painting' 등의 도장 관련 작업을 거쳐 'HEAVY+PE Area'에서 탑재가 되는 것을 확인할 수 있다. 조립 작업 후에 적치장인 'Stock Area'를 거쳐 탑재 되는 경우, 의장 작업 후에 적치장을 거쳐 탑재 되는 경우 등을 파악할 수 있다. 또한 사례 프로젝트의 경우, 의장 작업 후에 안벽 주변 적치장에서 추가 작업을 하거나 적치된 후, 다음 작업으로 이동하는 것을 알 수 있다. 뿐만 아니라, 도장 관련 작업의 경우는 'Blasting Shop', 'Painting Shop', 'Painting Area(Outdoor)'의 순서로 이동 프로세스가 이루어지므로, 적치장을 거치지 않고 순차적으로 작업이 비교적 잘 이루어진 인스턴스들로 군집 0이 구성되었다는 것을 알 수 있다. 군집 1의 <Figure 5(b)>와 군집 2의 <Figure 5(c)>도 단계 2에서 확인 했던 군집의 프로세스 흐름 특성은 유지하면서, 현재 데이터 수준으로 모델이 표현되어

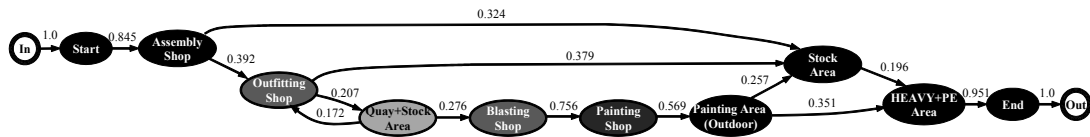


Figure 5(a). Standard Process Model of 'Cluster 0'

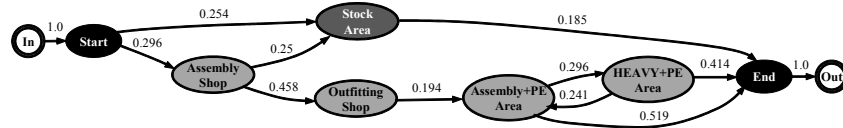


Figure 5(b). Standard Process Model of 'Cluster 1'

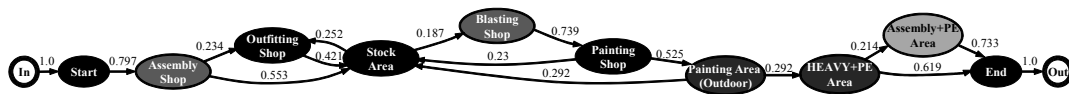


Figure 5(c). Standard Process Model of 'Cluster 2'

간단하고 명확해진 것을 확인할 수 있다.

각 군집별 표준 모델에 대한 적합도(fitness)는 식 (1)~식 (3)을 이용하여(Aalst, 2005) 각각 계산 하였다. 식 (1)의 경우는 케이스 즉, 프로세스 인스턴스 기준으로 계산하는 것이고, 식 (2)와 (3)은 이벤트를 기준으로 적합도를 계산하는 방법이다. 식 (2)와 식 (3)은 식은 동일하지만 이벤트가 실행되지 않았을 때, 블로킹(blocking)이 발생했을 때 처리하는 방법에 차이가 있다. 적합도 2(Fitness 2)의 경우는 특정 이벤트가 실행되지 않을 때, 블로킹(blocking)이 발생하여 이후 이벤트도 모두 실행될 수 없다고 판단하는 것이고, 적합도 3(Fitness 3)의 경우는 프로세스 내에 특정 이벤트가 실행되지 못해도, 이후 다른 이벤트는 조건에 따라 실행될 수 있다고 간주하고 분석하는 것이다.

본 대상 문제와 같이 이벤트의 발생이 매우 복잡하고, 이벤트 발생 위치와 상관없이 특정 시퀀스의 패턴을 확인하고 모델을 평가하고자 한다면 적합도 3이 가장 적합한 방법이라고 하겠다. 만약 본 사례연구에서 적합도가 더 높은 표준 모델을 얻고자 한다면, 제안한 프레임워크 내에서 로그 군집의 개수를 늘이면 된다. 예를 들어, 현재 제안한 로그 군집 내에서 추가적인 군집화를 수행하고 프로세스 모델을 얻을 수 있다. 그러나 이 경우는 보다 많은 노드와 노드 간 연결이 표준 프로세스 모델 내에 포함될 것이다. 따라서 표준 모델의 활용 목표를 정하고, 모델의 복잡도와 정확도 관점의 적정성을 고려하여 로그 군집의 개수를 결정하고 표준 모델의 수준을 결정해야 하겠다.

Table 2. Fitness for Standard Model of Each Cluster

	Fitness 1	Fitness 2	Fitness 3
Standard Model of Cluster 0	0.17	0.34	0.48
Standard Model of Cluster 1	0.20	0.26	0.31
Standard Model of Cluster 2	0.18	0.39	0.57

4.4 예외 프로세스 감지 (Step 4. Detecting exceptional processes)

각 군집별 정의한 표준 모델에 대해서 세 가지 유형의 적합도를 계산한 결과는 <Table 2>와 같다. 프로세스 인스턴스 관점에서 계산되는 적합도 1(Fitness 1)의 경우는 하나의 예외 이벤트만 발생해도 적합하지 않은 것으로 판단되기 때문에 적합도가 가장 낮다. 적합도 2의 경우는 프로세스 인스턴스 내에서 초반에 예외 이벤트가 발생하면 전체의 적합도를 더 많이 떨어뜨리기 때문에, 예외 이벤트의 발생위치에도 영향을 받는다.

단계 4에서는 단계 3에서 정의한 로그 군집별 표준 프로세스 모델 대비, 예외 이벤트와 프로세스를 감지하고 정리하는 작업을 한다. 생산 현장의 작업 관리자 혹은 프로세스 관리자 등은 표준 프로세스 모델을 정의하는 것에 만족하지 않고 예외 활동과 프로세스를 관리하기를 원한다. 군집별 표준 프로세스 모델을 이용하면 과거 실적 데이터에 대해서 모델과의 차이를 비교 분석할 수 있을 뿐만 아니라, 실시간 모니터링이 가능하면 표준 모델에서 벗어나는 예외 상황을 바로 감지할 수 있다.



**Table 3.** Frequency and Rate of Exceptional Events per Log Cluster

Case of Exceptional Events	Cluster 0 (total events : 1310)		Cluster 1 (total events : 556)		Cluster 2 (total events : 1301)	
	frequency	rate(%)	frequency	rate(%)	frequency	rate(%)
Case 1. Re-handling Movement	326	24.9	66	11.9	355	27.3
Case 2. Reverse Movement	90	6.9	11	2.0	36	2.8
Case 3. Unnecessary Movement	77	5.9	21	3.8	160	12.3

단계 3에서 정의한 예외 상황에 대해서 로그 군집별 발생 이벤트의 빈도(frequency)와 비율(rate)을 정리한 내용은 <Table 3>과 같다. <Table 3>에서 확인할 수 있는 것처럼, 로그 군집의 유형과 관계없이 Case 1의 ‘재 취급(Re-handling Movement)’이 가장 많은 것으로 파악된다. 재 취급은 데이터 상으로 동일 작업장이 순차적으로 연속해서 이동 로그로 남는 경우를 의미하는데, 이는 블록이 특정 작업에 바로 투입 되지 못하고 여러 번의 이동을 통해 투입된 경우 이거나, 작업 후에도 작업장을 바로 빠져 나오지 못하고 위치 이동이 있는 경우에 해당한다. 이러한 재 취급은 불필요한 이동과 경비를 발생시키므로 프로세스 개선을 통해 제거해 나가야 한다. Case 2의 ‘역 물류(Reverse Movement)’의 경우는 군집 1에서 가장 높은 비율로 발생했다. 군집 1은 주요 프로세스의 흐름을 가장 잘 따르면서, 모든 공정을 거쳐서 작업이 이루어지고 있어서 역 물류의 발생 빈도도 높은 것으로 판단된다. 상대적으로 군집 2와 군집 3은 프로세스 모델을 통해 이해되었던 것처럼, 특정 공정만 작업 후 탑재되는 흐름이 많아서 역 물류가 발생할 빈도가 낮은 것으로 분석된다. Case 3의 ‘불필요 이동(Unnecessary Movement)’의 경우는 군집 2의 경우가 비율이 가장 높다. 군집별 총 이벤트 수에 비해 인스턴스의 수가 가장 적기 때문에 그 만큼 블록별 이동회수가 많은 것을 의미하는데, 많은 이동 가운데 적치장간 이동도 많았던 것으로 분석된다. 해당 횟수만큼 블록들은 작업에 바로 투입되지 못하고, 적치장도 잘 확보하지 못한 것으로 이해할 수 있다.

## 5. 결론 및 토의

본 논문에서는 조선 산업에서 프로세스 마이닝을 이용한 블록 이동 프로세스 분석을 위한 프레임워크를 제안 하였다. 데이터 전처리 작업을 제외하고 목적에 따라 네 단계로 구성된다. 단계 1은 시스템의 로그 데이터로부터 전체 프로세스의 개요를 이해하는 것을 목적으로 하고, 단계 2는 관리할 수 있는 프로세스 모델을 얻기 위해서 로그를 군집화 하는 것을 목표로 한다. 단계 3은 실제 프로세스 관리에 활용할 수 있는 표준 프로세스 모델을 로그 군집별로 얻기 위한 작업으로 구성되고, 단계 4는 정의한 표준 모델과 대비하여 예외 프로세스를 감지하고 해당 내용을 레포팅 하는 것을 목적으로 한다. 제안한 프

레이워크에서는 프로세스 마이닝 기법들 중, 단계 1에서 퍼지 마이닝을, 단계 2에서 자취 군집화 기법을 사용하였다. 뿐만 아니라, 데이터의 의미론적 표현을 조정하는 방안, 퍼지 마이닝의 추상화 결과를 활용하는 방안 등을 제안 하였고, 단계 3과 단계 4에서는 본 논문의 대상 문제에 맞는 표준 프로세스 모델을 정의하는 방법과 예외 활동 및 프로세스를 분석하는 기준을 제시 하였다. 제안한 분석 프레임워크는 실제 조선소의 블록 공정 모니터링 시스템의 로그 데이터를 활용한 사례 연구를 통해 효용성을 제시하였다.

블록 공정 모니터링 시스템은 블록의 현재 위치 정보를 모니터링 할 목적으로 개발 되었고, 안정화되어 데이터의 신뢰도가 높다. 그러나 본 연구에서는 위치 정보 관점에서는 필요 없는 데이터일 수 있지만, 해당 블록의 이동이 어떤 목적으로 발생 했는지에 대한 부가적인 정보가 없는 것이 아쉬웠다. 조선 산업의 특성상, 특정 작업장에 블록이 위치했다고 해서, 해당 작업 중인 것으로 이해할 수 없는 경우도 있기 때문이다. 대표적으로 선행탑재 작업장의 경우, 도크 주변의 빈 공간이므로 실제 선행탑재 작업을 위해 위치하겠지만, 경우에 따라서는 적치장이 부족하여 잠시 블록을 적치해 둘 수도 있다. 특정 작업을 위한 주요 적치장도 꼭 해당 작업을 위해서만 적치 공간으로 활용된다는 것을 보장할 수는 없다. 단순히 블록의 위치 정보만으로는 일부 이러한 예외 상황들을 구분해 내기에는 어려움이 있어서, 예외 상황에 대한 구분에 전문가의 판단이 필요했다. 따라서 작업자의 입력을 추가하는 일이지는 하지만 블록 공정 모니터링 시스템에 블록 이동 목적 정보도 추가 관리하면 보다 정확한 분석이 가능할 것이다.

작업장이 넓고, 작업자가 많고, 중간 제품(블록)의 크기가 크고, 부품이 많은 조선 산업의 특성상 생산부서의 상세 작업 프로세스를 이해하고 관리하기가 힘들다. 본 연구와 같이 데이터를 통해 생산의 프로세스를 조망하는 것은 사람과 문건을 통해 이해하는 것보다 훨씬 빠르고, 정확하게 많은 이해를 가능하게 한다. 뿐만 아니라, 이러한 분석은 프로세스 개선과 많은 의사결정에 활용될 수 있을 것으로 사료된다.

향후 연구로는 주요 작업 공정에 대한 일정 계획 정보와 실적 기준의 블록 이동 프로세스 모델을 비교 분석 하는 것과 정의한 표준 프로세스 모델을 이용해서 예외 프로세스들을 관리하고, 실제 프로세스를 평가하는 구체적인 방법론 등을 고려하고 있다.

## 참고문헌

- van der Aalst, W. M. P. and Basten, T. (2002), Inheritance of workflows : An approach to tackling problems related to change, *Theoretical Computer Science*, **270**(1), 125-203.
- van der Aalst, W. M. P. (2005), Business alignment: using process mining as a tool for Delta analysis and conformance testing, *Requirement Engineering*, **10**(3), 198-211.
- van der Aalst, W. M. P., Reijers, H. A., and Song, M. (2005), Discovering social networks from event logs, *Computer Supported Cooperative work*, **14**(6), 549-593.
- van der Aalst, W. M. P., van Dongen B. F., Günther, C. W., Rozinat, A., Verbeek, E., and Weijters, T. (2009), ProM : the process mining toolkit, BPM Demonstration, *CEUR Workshop Proceedings*, **489**, 1-4.
- Berry, M. J. A. and Linoff, G. (1997), *Data Mining Techniques*, John Wiley and Sons, Inc.
- Günther, C. W. and van der Aalst, W. M. P. (2007), Fuzzy mining -Adaptive process simplification based on multi-perspective metrics, BPM Workshops, *Lecture Notes in Computer Science*, **4714**, 328-343.
- Jansen-Vullers, M. H., van der Aalst, W. M. P., and Rosemann, M. (2006), Mining configurable enterprise information systems, *Data and Knowledge Engineering*, **56**(3), 195-244.
- Lee, S., Kim, J., and Moon, I. (2011), Deployment planning of blocks from storage yards using a tabu search algorithm, *Journal of the korean institute of industrial engineers*, **37**(3), 198-208.
- Lee, S., Kim, B., Huh, M., Cho, S., Park, S., and Lee, D. (2013), Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry, *Expert Systems with Applications*, **40**(1), 83-95.
- de Medeiros, A. K. Alves, Weijters, A. J. M. M., and van der Aalst, W. M. P. (2007), Genetic Process Mining : An Experimental Evaluation, *Data Mining and Knowledge Discovery*, **14**(2), 245-304.
- Park, C. and Seo, J. (2012), A GRASP approach to transporter scheduling and routing at a shipyard, *Computer and Industrial Engineering*, **63**, 390-399.
- Rozinat, A. and van der Aalst, W. M. P. (2006), Decision Mining in ProM, Proc. 4th Int. Conf. on Business Process Management, 420-425.
- Song, M., Günther C. W., and van der Aalst, W. M. P. (2009), Trace clustering in process mining, BPM 2008 Workshops, *Lecture Notes in Business Information Processing*, **17**, 109-120.
- Veiga, G. M. and Ferreira, D. R. (2010), Understanding Spaghetti Models with Sequence Clustering for ProM, BPM 2010 Workshops, *Lecture Notes in Business Information Processing*, **43**, 92-103.
- de Weerd, J., de Backer, M., Vanthienen, J., and Baesens, B. (2012), A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs, *Information Systems*, **37**, 654-676.
- Yahya, B. N., Park, J. H., Bae, H. and Mo, J. K. (2011), Similarity measurement using ontology in vessel clearance process, *Journal of the korean institute of industrial engineers*, **37**(2), 153-162.
- Yu, Y.-W., Kim, S., and Bae, H. (2012), Business process modeling using process structure constraints and social relations, *IE Interfaces*, **25**(3), 300-308.