http://www.jsebs.org
ISSN: 2288-3908

The Journal of Society for e-Business Studies
Vol.23, No.2, May 2018, pp.97-110
https://doi.org/10.7838/jsebs.2018.23.2.097

스마트제조를 위한 머신러닝 기반의 설비 오류 발생 패턴 도출 프레임워크

A Machine Learning Based Facility Error Pattern Extraction Framework for Smart Manufacturing

윤준서(Joonseo Yun)*, 안현태(Hyeontae An)**, 최예림(Yerim Choi)***

초 록

4차 산업혁명 시대를 맞아, 제조 기업들은 생산성 향상을 위해 축적된 설비 데이터를 활용하여 스마트제조를 실현하는 것에 높은 관심을 두고 있다. 하지만 기존의 설비 데이터 분석 연구들은 주로 센서 데이터 등 정형 데이터를 대상으로 하여, 실제 큰 비중을 차지하고 있는 텍스트와 같은 비정형 데이터에 대한 분석 연구는 부족한 실정이다. 특히, 작업자가 수기로 작성한 텍스트 데이터를 활용한 사례는 매우 적었다. 따라서 본 논문에서는 작업자가 수기로 작성한 설비 오류 데이터를 분석하여 연관 규칙 마이닝을 통해 설비 오류 발생 패턴을 도출하는 프레임워크를 제안하고자 한다. 이때, 일반적인 텍스트 분석 기법과 같이 단어를 분석 기준으로 사용하는 경우 전문 용어에 해당하는 설비 오류의 의미를 표현하는 데에 한계가 있다는 점에 착안하여 구절을 추출하여 텍스트 분석 기준으로 사용하였다. 제안하는 프레임워크의 성능을 실제 사례를 통해 검증하였으며, 본 연구 결과를 활용하면 설비 오류를 예방하여 가동률을 높이고 나아가 제조 기업의 생산성 향상에 기여할 수 있을 것으로 기대한다.

ABSTRACT

With the advent of the 4-th industrial revolution, manufacturing companies have increasing interests in the realization of smart manufacturing by utilizing their accumulated facilities data. However, most previous research dealt with the structured data such as sensor signals, and only a little focused on the unstructured data such as text, which actually comprises a large portion of the accumulated data. Therefore, we propose an association rule mining based facility error pattern extraction framework, where text data written by operators are analyzed. Specifically, phrases were extracted and utilized as a unit for text data analysis since a word, which normally used as a unit for text data analysis, is unable to deliver the technical meanings of facility errors. Performances of the proposed framework were evaluated by addressing a real-world case, and it is expected that the productivity of manufacturing companies will be enhanced by adopting the proposed framework.

키워드 : 스마트제조, 설비 오류 패턴 도출, 머신러닝, 비정형 데이터, 텍스트 데이터 분석, 연관 규칙 마이닝

Smart Manufacturing, Facility Error Pattern Extraction, Machine Learning, Unstructured Data, Text Data Analysis, Association Rule Mining

본 연구는 경기도의 경기도 지역협력연구센터 사업의 일환으로 수행하였음[GRRC경기2017-B01, 지능형 제조 빅데이터 분석 연구].

* First Author, Department of Industrial and Management Engineering, Kyonggi University(jsyun2009@gmail.com)
** Co-Author, Department of Industrial and Management Engineering, Kyonggi University(hyeontaean94@gmail.com)
*** Corresponding Author, Department of Industrial and Management Engineering, Kyonggi University (yrchoi@kgu.ac.kr)

Received: 2018-04-10, Review completed: 2018-05-15, Accepted: 2018-05-21

1. 서 론

최근 제조 분야에서는 4차 산업혁명에 관한 관심이 증가하고 있다[7, 17]. 4차 산업혁명은 인공지능, 빅데이터, 사물인터넷 등 첨단 ICT 기술이 기존의 제조 기술들과 융합되어 생산성을 향상시키는 차세대 산업혁명을 의미하며[27], 4차 산업혁명을 견인하는 기술 중 대표적인 빅데이터 기술은 사물인터넷을 활용하여 많은 양의 데이터를 수집하고 발달된 컴퓨터의 처리 능력을 활용하여 데이터로부터 가치를 창출하는 기술을 의미한다[5, 24]. 한국을 포함한 주요 선진국들에서는 급변하는 시장 상황을 극복하기 위해서는 제조 분야에서 빅데이터를 활용한 새로운 가치를 창출이 필요하다고 인식하고 있으며[13, 22], 이에 따라 빅데이터 기술을 활용한 스마트제조(smart manufacturing) 관련 연구가 활발히 진행되고 있다[21, 23].

빅데이터 분석을 위해서는 많은 양의 데이터를 수집이 선행되어야 한다. 과거 제조 현장에서는 작업자의 실수나 열악한 환경으로 인해 데이터의 정보가 누락되거나 실시간이 결여되는 경우가 발생하기 때문에 데이터에 대한 신뢰성이 떨어졌다[18]. 반면 최근 스마트공장에서는 사물인터넷 기술의 발달로 다양(variety)하고 많은 양(value)의 데이터가 빠른 속도(velocity)로 수집되고 있어[6, 15, 25], 제조 빅데이터를 이용한 분석이 가능해졌다.

제조 설비에서 수집된 데이터는 정형화 정도에 따라 크게 센서 신호와 같은 정형 데이터와 텍스트, 이미지, 동영상 등의 비정형 데이터로 구분할 수 있다[12, 24]. 아직까지 제조 현장에서는 데이터 수집의 자동화가 불가능한 부분이 존재하여, 데이터를 기록하기 위해 작업자들이 직접 수기로 기록을 남기는 경우가 빈번하게 발생한다. 따라서, 비정형 데이터의 일종인 텍스트 데이터가 제조 현장에서 수집되는 데이터의 많은 부분을 차지하고 있다[1].

하지만 제조 설비에서 수집되는 데이터를 활용한 기존 연구들은 주로 정형 데이터를 활용한 경우가 많으며, 비정형 데이터를 활용한 연구는 상대적으로 적었다. 정형화 정도에 따라 기존 제조 데이터 분석 연구를 살펴보면 다음과 같다. 변성규 외[3]의 연구에서는 자동차 공장의 온도와 같은 정형 데이터를 활용하여 공정 단계에서 미리 불량 패턴과 유형을 파악하여 불량을 사전에 예측하였으며, 이우배 외[19]에서는 산업단지의 성장에 영향을 미칠 수 있는 다양한 변수를 정형 데이터로부터 도출하여 산업단지의 성장이 고용 창출에 미치는 영향을 분석하였다. 또한, 설비의 비정상 상태를 탐지하기 위해 터키의 설탕 공장으로부터 수집한 정형 데이터를 분석한 연구[28]나 현장에서 수집된 데이터의 평균, 표준편차 등을 분석한 연구[29]가 있었다. 드물게는 반도체 공정의 수율을 개선하고 예측하기 위해 비정형 데이터를 활용하는 연구가 있었다[2].

일찍이 제조 분야에서는 생산성을 개선하기 위한 다양한 노력이 있었다. 제조 생산성은 노동 생산성과 설비 생산성으로 나뉘며, 이 중 설비 생산성은 설비 효율화의 양적 측면에서 가동률과 연관이 있다. 특히, 제조 설비의 경우 대부분이 고가의 장비들이기 때문에 제조업체는 주로 설비의 가동률을 높여 생산성을 향상시키는 방법으로 원가 우위는 점한다[29]. 설비의 가동률은 전체 설비에서 가동 가능한 설비의 수를 의미한다[26]. 즉, 제조업에서 생산성을 높이기 위해서는 설비의 가동률을 낮추는 요인

을 예방해야 하며[10, 28], 이를 위해서는 제조 설비에서 수집되는 데이터를 분석하는 빅데이터 기술이 활용될 수 있다.

본 연구에서는 제조 현장에서 많은 부분을 차지하고 구체적인 정보를 담고 있는 비정형 텍스트 데이터를 분석하여 의미 있는 결과를 찾기 위한 연구를 수행하고자 한다. 구체적으로, 제조 생산성 향상을 위해 설비 가동률을 낮추는 설비 오류의 발생 패턴을 도출하여 이를 예방하고자 한다. 설비 오류 발생 패턴을 도출하기 위해서는 과거의 설비 오류 이력 데이터가 필요하며, 이러한 데이터는 일반적으로 텍스트 데이터를 포함한 비정형 데이터이다. 예를 들어, 본 연구에서 사용한 설비 오류 이력의 경우 작업자가 설비 오류의 원인이나 현상, 조치를 수기로 작성한 비정형 데이터이다.

텍스트 데이터의 경우 인간의 대표적인 정보 전달 수단으로, 소셜미디어의 발전과 함께 대량 수집이 용이해지며 텍스트 데이터 분석을 주제로 활발한 연구가 이뤄지고 있다. 하지만 제조 설비에서 생성된 데이터를 다루는 연구는 극히 드물고, 대부분은 제품 사용 후기나 불만사항, 시장 동향/전망 정보, 소셜미디어와 같이 고객으로부터 발생하는 데이터를 다루었다[8, 9, 20, 22]. 일반적인 텍스트 데이터 분석 기법들은 분석 단위로 하나의 단어를 이용한다. 예를 들어, 최근 주목받고 있는 텍스트 분석 기법의 하나인 word2vec[11]은 단어 의미 유사성을 기반으로 하나의 단어를 벡터 공간에 할당하여 관련성을 표현하는 방식을 사용한다. 하지만 전문 용어가 많은 설비 오류 이력 데이터의 경우 하나의 단어가 독립적으로 사용되면 전달하고자 하는 의미가 달라질 수 있다는 문제점이 있다. 예를 들어, '부품 파손'과 '부품 노후'는 '부품'이라는 단어가 동시에 등장하지만, 별개의 의미를 전달하며, '부품 파손'과 '스프링 파손' 또한 의미적으로 구분되어야 한다.

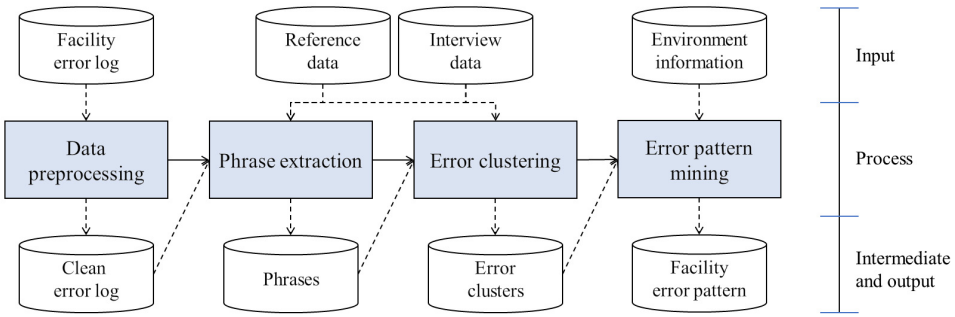
따라서 본 연구에서는 작업자가 수기로 작성한 비정형 텍스트 데이터를 분석하여 설비의 오류 발생 패턴을 도출하는 프레임워크를 제안한다. 특히, 하나의 단어를 분석 단위로 사용하는 대신 단어의 나열인 구절을 기준으로 텍스트 데이터 분석을 수행하여 하나의 구절이 의미를 전달하는 전문 용어의 특징을 반영하였다. 제안하는 프레임워크는 크게 데이터 전처리, 구절 추출, 오류 군집화, 오류 패턴 마이닝의 네 단계로 구성된다.

본 논문은 다음과 같은 순서로 구성된다. 제 2장에서는 본 논문에서 제안하는 오류 발생 패턴 추출 프레임워크를 자세히 서술한다. 제 3장에서는 제안하는 프레임워크를 실제 사례에 도입하는 과정을 제시하고 그 결과를 해석한다. 마지막으로 제 4장에서는 연구 내용을 정리하고 향후 연구 방향 및 기대 효과를 논의한다.

2. 설비 오류 발생 패턴 도출 프레임워크

2.1 개요

본 연구에서 제안하는 설비 오류 발생 패턴 도출 프레임워크는 크게 데이터 전처리, 구절 추출, 오류 군집화, 오류 패턴 마이닝의 네 가지 단계로 진행된다. <Figure 1>은 제안하는 프레임워크의 전체 구조를 나타낸다. 설비 오류 이력, 참고 문헌, 인터뷰 데이터, 기상 정보를 입력받아 데이터 전처리, 구절 추출, 오류 군집화,



〈Figure 1〉 Overview of the Proposed Facility Error Pattern Extraction Framework

오류 패턴 마이닝의 네 가지 단계를 진행하여 최종적으로 설비 오류 패턴을 도출한다.

데이터 전처리 단계에서는 제조 공정에서 발생하는 비정형 데이터의 하나인 설비 오류 이력을 입력받아 오탈자 등 노이즈를 제거하고 텍스트 데이터 분석 기법을 도입하기에 적합한 형태로 변환한다. 다음으로, 구절 추출 단계에서는 참고 문헌과 현장 전문가와의 인터뷰 데이터를 토대로 작업자가 작성한 문장 형태의 오류 이력 텍스트 데이터를 의미 전달이 가능한 구절 단위로 파싱한다. 오류 군집화 단계에서는 구절 단위로 쪼개진 오류 발생 이력을 소수 개의 오류 유형으로 군집화한다. 마지막으로, 오류 패턴 마이닝 단계에서는 날씨 정보 등 환경 정보를 추가적으로 입력받아 대표적인 연관 규칙 마이닝 기법의 하나인 FP-Growth(frequent pattern growth) 알고리즘을 이용하여 오류 발생 패턴을 도출한다.

2.2 데이터 전처리

본격적인 텍스트 데이터 분석에 앞서 가장 먼저 설비 오류 이력 데이터를 분석 가능한 형태로 정리한다. 데이터 전처리는 두 가지 단계로 진행된다. 먼저, 전체 오류 이력에서 분석에 불필요한 부분을 제거하고, 남은 데이터에서 오탈자 등 오류를 수정한다.

설비에서 발생한 오류 상황을 기준으로 수집된 설비 오류 이력 데이터는 설비 코드나 고장 원인, 오류 시작일 등과 같은 다양한 항목으로 구성될 수 있다. 본 연구에서 중점적으로 활용하는 데이터는 작업자가 설비 오류 상황에 대해서 수기로 작성한 비정형 텍스트 데이터이므로 이외의 요인은 제거하여 사용한다. 또한, 설비 오류 이력 데이터에서 설비 오류 상황으로 판단하기 어려운 데이터는 제외한다. 예를 들어, 자재 품절과 같은 상황은 본 연구에서 다루고자 하는 오류의 범주를 벗어난다고 판단할 수 있어 본 단계에서 제거된다.

제안 프레임워크에서 사용할 데이터가 일차적으로 분리되면, 이를 대상으로 오탈자 등 오류를 수정한다. 구체적으로, 맞춤법이 틀린 단어들을 수정하고, 동일한 의미를 가지지만 작업자에 따라 다양한 방식으로 기입된 단어들은 한 단어로 통일한다. 예를 들어, 외래어인 컨베이어의 경우 작업자에 따라 '컨베어', '컨배이어', '컨베이어'로 다르게 표기될 수 있어 분석에 앞서 이러한 단어들을 통일된 표현으로 정리한다. 또한, 문장의 의미를 알 수 없는 데이터는 제외한다.

2.3 구절 추출

전처리를 마친 설비 오류 텍스트 데이터를 기반으로 구절을 추출한다. 여기서 구절이란 연결된 단어들의 리스트로 설비 오류에 대한 단일의 의미를 가진 최소한의 단위로 이후 단계에서 분석 단위로 사용된다. 구절 추출 단계에서는 전 처리된 데이터 외에도 구절 단위로 파싱하는 과정에서 참고 문헌과 제조 현장 전문가와의 인터뷰를 통해 관련 지식을 수집하여 구절 추출에 반영한다.

본 연구에서의 구절 추출은 일반적인 사전 기반의 단어 추출 방식을 동일하게 활용한다. 사전 기반 단어 추출의 경우 가능한 단어의 리스트를 사전 형태로 구축해두고 이를 이용하여 일반 텍스트에서 사전과 매칭되는 단어를 추출한다. 따라서, 본 단계에서 가장 먼저 설비 오류에 대한 구절 사전을 구축한다. 특히, 참고 문헌과 현장 전문가와의 인터뷰 데이터를 활용하여 누락되거나 잘못된 구절이 없도록 사전을 구축한다. 또한, 현재 텍스트 데이터에는 포함되어 있지 않지만 추후 포함될 수 있거나 포함되어야 한다고 생각되는 구절로 변환하여 추가한다.

위와 같이 구축된 설비 오류 구절 사전을 이용하여 전체 설비 오류 이력 데이터를 구절화시킨다. 구체적으로, 설비 오류 이력 데이터에서 작업자가 수기로 작성한 텍스트 부분을 설비 오류 구절 사전과 매칭하여 각각의 텍스트로부터 오류와 관련된 구절을 추출한다.

2.4 오류 군집화

본 단계에서는 제 2.3절에서 추출된 구절을 군집화한다. 군집화를 수행하는 이유는 더욱 정확한 오류 패턴 마이닝을 수행하기 위해서이다. 오류 패턴 마이닝을 통해 오류가 어떤 상황에서 발생하는지 파악하기 위해서는 상황을 묘사하는 부분과 오류 타입의 쌍으로 구성된 데이터를 사용하며, 여기서 오류 타입 별 상황에 대한 데이터가 충분해야지만 유의미한 결과를 얻을 수 있다. 따라서 다양한 표현으로 기재된 오류 구절을 군집화를 수행하여 소수 개의 오류 타입으로 변환한다. 군집화 단계에서도 설비 오류와 관련된 참과 문헌과 제조 현장 전문가와의 인터뷰를 통해 얻은 지식을 반영한다.

오류 군집화는 총 두 단계에 걸쳐서 진행되었다. 첫 번째 단계는 표현은 다르지만 동일한 의미를 가지는 구절들끼리 군집화하여 1단계 오류 유형 군집을 만들었다. 두 번째 단계에서는 1단계 오류 유형 군집 중 유사하다고 판단되는 군집들끼리 묶어 2단계 오류 유형 군집을 만들었다. 이때, 유사하다는 의미는 구절이 서로 의미적으로 유사하고 판단되거나 동일한 조치 내역을 필요로 한다고 판단되는 것을 의미한다. 이후 단계에서는 2단계 오류 유형 군집을 오류 타입으로 지정하여 사용한다.

2.5 오류 패턴 마이닝

본 연구에서는 연관 규칙 마이닝의 대표적인 방법론인 FP-Growth를 활용하여 설비 오류 발생 패턴을 파악한다. FP-Growth는 자주 일어나는 반복적인 패턴을 찾아내는 알고리즘으로[4], 선행 규칙(원인)과 결과 규칙(결과)으로 구성된 데이터를 분석하여 이들의 조합이 나타나는 횟수와 확률을 알 수 있다. 구체적으로, 선행 규칙과 결과 규칙의 조합에 대해 특정 조합의 발생 빈도를 의미하는 support 값과 특정

조합의 선행 규칙이 만족하였을 때 결과 규칙이 발생할 가능성인 confidence 값을 제공한다.

Support 값은 선행 규칙과 결과 규칙이 함께 표현된 횟수를 나타내며, 식 (1)과 같이 계산한다.

$$\text{support}(A \rightarrow B) = P(A \cap B) \quad (1)$$

이때, A는 선행 규칙, B는 결과 규칙을 의미한다. Confidence 값은 support 값을 선행 규칙이 나타나는 전체 경우의 수 나눈 값으로, 식 (2)와 같이 계산한다.

$$\text{confidence}(A \rightarrow B) = \frac{P(A \cap B)}{P(A)} \quad (2)$$

설비 오류 발생 패턴 도출을 위해 사용될 수 있는 선행 규칙으로는 설비 오류가 발생한 시점에 대한 정보(월, 요일, 시간 등), 오류가 발생한 위치 대한 정보(설비명, 고장개소그룹 등), 기타 설비 오류 발생 시점에 주위 환경 정보(작업자, 날씨 등) 등이 있다. 결과 규칙으로는 제 2.4절에서 생성된 2단계 오류 유형이 사용된다. 이와 같은 선행 규칙과 결과 규칙으로 구성된 데이터를 구축하기 위하여 정제된 오류 이력 데이터, 군집화 결과, 외부로부터 수집한 환경 정보를 매칭한다.

3. 사례 연구

본 연구에서 제안하는 설비 오류 발생 패턴 도출 프레임워크를 평가하기 위해 실제 제조 기업에서 수집된 데이터를 이용하여 설비 오류 발생 패턴을 도출하였다. 단계별 진행 내용과 결과물은 아래에서 상세히 설명한다.

3.1 사용 데이터

본 사례에서는 제안 프레임워크를 진행하기 위해 크게 네 종류의 데이터를 수집 및 사용하였다. 분석을 위해 제조 기업에서 축적한 설비 오류 이력 데이터를 사용하였으며, 추가적으로 배경이론의 축적을 위해 현장의 작업자와 인터뷰를 통해 수집된 현장 데이터와 참고 문헌 데이터를 사용하였다. 또한, 오류 발생 패턴의 신뢰성을 높이기 위해 선행 규칙에서 사용될 환경 정보로 공장이 위치한 지역의 기상 데이터를 사용하였다.

본 연구에서 사용된 설비 오류 이력 데이터는 세 가지 공정으로부터 약 사 년 동안 수집된 것으로, <Table 1>과 같이 설비 코드, 설비명, 고장 현상, 고장 원인, 고장개소그룹, 고장 개소, 조치내역, 조치자, 시작일, 완료일의 10가지 항목으로 구성되어 있다. 현장 정보 수집을 위한 작업자와의 인터뷰는 총 두 차례 실시하였다. 1차 인터뷰에서는 현장에서 통용되는 전문 용어들이나 오류 이력에서 의미가 명확하지 않았던 표현들에 관한 질의응답이 진행되었다. 2차 인터뷰에서는 기입된 내용의 오류로 분석에서 제외되는 데이터를 줄이기 위해 비슷한 고장 원인이나 현상과 대조하여 활용 가능한 데이터 수를 증가시켰다. 참고 문헌 데이터로는 제조 설비의 상태 진단 항목[16], 설비 작업의 인적 오류 평가 방법[30], 오류 방지 매뉴얼[14] 등을 이용하였다. 마지막으로, 기상 데이터는 기상자료개방포털에서 제공하는 14년에서 17년까지의 일시, 기온, 강수량, 습도, 현지 기압, 적설의 6가지 항목으로 구성된 자료를 활용하였다.

〈Table 1〉 Examples of Facility Error Log

Facility code	Facility name	Error phenomenon	Cause of Error	Error location	Error site	Facility error text	Operator	Start date	End date
10091	OZ Gasket Neversize Applicator#3	Malfunction	Unknown cause	AIR equipment	Pipe	Since the application amount was not constant, the input value of the application part controller was changed	A	14.05.14 8:36	14.05.22 13:17
10091	OZ Gasket Neversize Applicator#3	Malfunction	Component aging	Electric equipment	Sensor	Phenomenon: An abnormal phenomenon occurred in the application part. Cause: The product presence sensor of gasket detector did not operate smoothly. Action: The product presence sensor was replaced	A	14.08.19 15:47	14.08.19 15:50

데이터 전처리 단계에서는 가장 먼저 세 가지 공정에서 수집된 설비 정지 이력 데이터를 하나의 파일로 통합하였다. 다음으로 텍스트 분석을 수행하기 위해 고장 현상, 고장 원인, 조치 내역의 세 가지 항목을 하나의 항목으로 합쳤다. 마지막으로 잘못된 데이터를 삭제 혹은 수정하였다. 고장 현상이 빈칸인 경우, 문장의 의미를 알 수 없는 경우, 자재 품절과 같이 오류로 볼 수 없는 경우는 분석에서 제외하였다. '컨베어', '컨배이어', '컨베이어'와 같이 작업자에 따라 다르게 표기한 데이터는 하나로 통일하였다.

3.2 구절 추출 결과

구절 추출 단계에서는 오류 이력에 있는 텍스트 데이터로부터 유의미한 구절을 추출하는 단계이다. 가장 먼저 의미를 가진 가장 최소한의 단위인 중요 구절로 구성된 구절 사전을 구축하고, 구절 사전에 참고 문헌과 인터뷰 내용을 반영한 중요 구절들을 추가하였다. 구절 사전을 바탕으로 설비 오류 이력에서 작업자가 수기로 작성한 항목을 구절화하고 매칭하여 설비 코드, 설비명, 시작일 등의 10가지 항목으로 구성된 데이터를 만든다.

〈Table 2〉 Examples of the Facility Error Texts Utilized for Phrase Extraction

Number	Facility error text
1	overload/overload/2 conveyor overload, center adjustment after action
…	……
569	malfunction/unknown cause/2 times rim cycle fault, stop, frequent occurrence, cycle time fit sensor height detection value adjustment
…	…
1311	quality abnormality/quality abnormality/symptom: heater overload, cause: wiring No.3, action: wiring No.3, temperature measurement

<Table 2>는 설비 오류 이력에서 뽑힌 텍스트 데이터를 나타낸다. 이러한 데이터로부터 미리 구축된 구절 사전과의 매칭을 통해 구절을 추출한다. <Table 3>은 구절 추출 단계를 통해 얻은 구절들의 예시를 보여준다. 설비 오류 이력은 총 1,311개의 행으로 구성되어 있으나 하나의 오류 이력에서 다수개의 구절이 추출될 수 있어서 오류 구절은 총 1,593개가 추출되었다.

〈Table 3〉 Examples of the Phrases Extracted from Facility Error Texts

Number	Phrase
1	2 conveyor overload
2	center adjustment
3	overload
…	…
706	2 abnormal cycle of rim cycle
707	cycle time alignment
708	malfunction
…	…
1589	above quality
1590	heater overload
1591	3 wiring error
1592	reconnect wiring 3
1593	temperature measurement

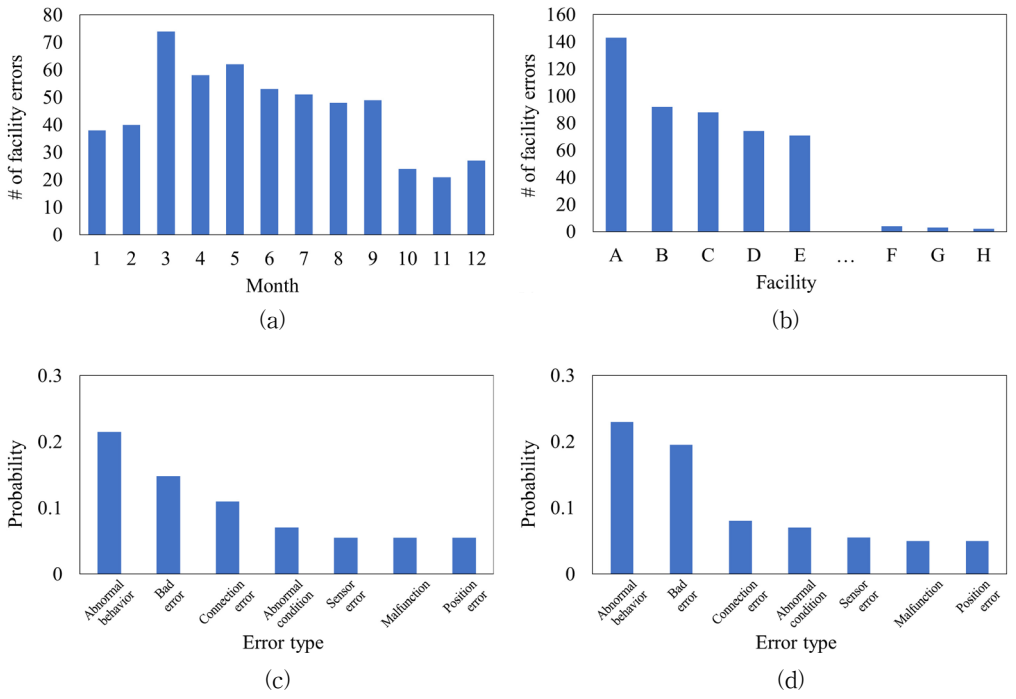
3.3 오류 군집화 결과

군집화는 총 2단계에 걸쳐서 진행되었다. 1단계는 구축한 사전에서 표현은 다르지만 동일한 의미를 가지는 중요 구절들끼리 군집화하여 1단계 오류 유형 군집을 만들었다. 2단계에서는 유사하다고 판단되는 1단계 레이블을 정성적으로 군집화하여 2단계 오류 유형 군집을 만들었다. <Table 4>는 오류 유형 군집화 결과 얻어진 2단계 오류 유형 군집을 나타낸다. 1단계에서는 총 백 개 이상의 군집이 도출되었으며 2단계에서는 총 30개의 오류 유형 군집이 도출되었다.

본 연구에서는 2단계 오류 유형을 오류 타입으로 사용하였다. <Figure 2>는 오류 타입의 특징을 가시화한 결과이다. (a)를 통해 3월에 오류 발생이 가장 많으며, 11월에 가장 적음을 확인할 수 있다. 특히, (c)는 3월에 발생하는 오류 타입을 빈도순으로 정리한 그래프로, '이상동작'이 가장 높은 빈도를 차지하는 것을 확인할 수 있다. (b)는 설비별 오류 발생 횟수로, 가장 높은 빈도수를 보인 설비 A도 '이상동작' 오류 타입의 비중이 가장 높음을 (d)를 통해 확인할 수 있다.

〈Table 4〉 Full List of the Error Types Extracted from the Case Study

Center error	Position error	Above slope	Pressure error	Air abnormal	Leakage
Aging	Abnormal cylinder	Open-close error	Sensor error	Alarm error	Noise
Temperature abnormality	Cycle over	Size	Forward and backward	Over speed	Not supplied
Foreign matter	Malfunction	Computer equipment error	Impossible	Measurement change	Bad error
Not working	Abnormal condition	Connection error	Abnormal behavior	Human error	Component error



〈Figure 2〉 Overview of Facility Errors In Terms of (A) The Number of Facility Errors Per Month, (B) The Number of Facility Errors Per Facility, (C) Probability of Error Occurrences Per Error Type for March, and (D) Probability of Error Occurrences Per Error Type for Facility A

3.4 오류 패턴 마이닝 결과

오류 패턴 마이닝 단계에서는 가장 먼저 선행 규칙과 결과 규칙의 쌍으로 구성된 입력 데이터를 만든다. 우선, 구절 추출 단계에서 만들어진 설비 오류 이력의 구절과 군집화 단계에서 만들어진 2단계 군집을 매칭하였다. 다음으로 앞서 매칭된 데이터에 설비 오류 이력의 다른 구성 요소를 추가하여 시작일, 설비명, 고장 개소그룹, 조치자, 날씨, 구절, 2단계 오류 유형의 항목으로 구성된 입력 데이터를 만들었다. <Table 5>는 본 연구에서 사용한 입력 데이터를 나타낸다. 입력 데이터에 FP-Growth를 적용하여 오류 발생 패턴을 파악하였다.

오류 패턴 마이닝을 수행하여 다음과 같은 결과를 도출하였다. <Table 6>은 오류 패턴 중 비교적 높은 confidence 값을 갖는 경우의 예시이다. 결과를 해석해보면 선행 규칙이 '9월', 'Wed', '기계장치'일 때, 오류 타입 '동작 이상'이 5번 발생했고, 이때, confidence 값은 83.3%로 선행 규칙을 만족하는 오류 이력 중 '동작 이상'이 80% 이상을 차지함을 알 수 있다. 또한, 선행 규칙이 '8월', 'Thu', 'AIR 장치'이고 오류 타입이 '오작동'인 패턴과 선행 규칙이 '6월', 'OZ ASSY 성능검사기#2', '전기장치'이고 오류 타입이 '연결 이상'인 패턴은 각각 support 값은 4와 3이고, confidence 값이 100%로 선행 규칙을 만족하는 오류 이력은 모두 해당하는 오류 타입에 해당함을 알 수 있다.

〈Table 5〉 Input Data for Error Pattern Mining

Number	Month	Day of week	Weather	Facility name	Error location	Operator	Error type
1	May	Wed	Sunny	OZ ASSY Performance checker #2	Electric device	S	Bad error
2	May	Wed	Cloudy	OZ ASSY Performance checker #2	Electric device	S	Connection error
…	…	…	…	…	…	…	…
545	August	Tue	Sunny	OZ Straight type assembly machine #1	Machinery	C	Abnormal behavior

〈Table 6〉 Examples of the Extracted Error Patterns with High Confidences

Precedence rule	Result rule	Support	Confidence
[‘September’, ‘Wed’, ‘Mechanics’]	Abnormal behavior	5	0.83
[‘August’, ‘Thu’, ‘AIR device’]	Malfunction	4	1.00
[‘June’, ‘OZ ASSY Performance Checker #2’, ‘Electric Device’]	Connection error	3	1.00

4. 결 론

본 연구에서는 스마트제조 실현을 통한 제조업의 생산성 향상을 위해 제조 공정에서 발생하는 비정형의 텍스트 데이터를 이용하여 설비의 오류 발생 패턴을 도출하는 프레임워크를 제안하였다. 구체적으로, 설비 오류 이력의 특성상 한 단어로는 전문 용어의 의미를 표현하기가 어렵다는 점에 기반을 두어 구절을 기본 단위로 이용하여 설비 오류의 원인과 현상을 표현하였으며, 연관 규칙 마이닝의 대표적인 방법론인 FP-Growth 기법을 활용하여 오류 유형과 월, 요일, 설비명, 고장개소그룹, 조치자, 날씨 간의 관계를 확인하였다. 제안된 프레임워크를 통해 제조 공장 설비의 오류 유형을 파악하고 설비 정지를 예방하여 설비의 가동률을 높이고 나아가 생산성 향상에 기여할 수 있을 것으로 기대된다.

본 논문에서 제안하는 프레임워크는 아직 초기 단계로 주요 구절 추출 및 오류 군집화 과정은 정성적인 방법으로 이루어졌다. 따라서, 추후 연구에서는 제안된 프레임워크의 각 과정에 머신러닝 기법을 도입하여 고도화를 진행할 예정이다. 또한, 오류 발생 패턴 마이닝을 위해 추가적인 환경 요인들을 고려하여 더욱 의미있는 패턴을 찾을 수 있을 것으로 기대한다. 본 논문의 사례 연구에서 사용된 설비 데이터에는 오류 이력뿐만 아니라 각 오류에 대한 조치 내역이 포함되어 있다. 이러한 조치 내역에 대한 분석을 통해 설비 오류의 발생이 예측되었을 때 부품의 교체 등 요구되는 조치 방안을 함께 추천하는 기법 또한 개발될 수 있을 것이다.

References

[1] Bae, S. M., Lee, H. W., Lee, G. A., Choi, S., and Park, H. K., “Enhancing Manufacturing Data Quality for Data Mining,

Korean Society for Precision Engineering," In Proceedings of the Korean Society of Precision Engineering Conference, pp. 795-796, 2007.

[2] Baek, D. H. and Han, C. H., "Application of Data Mining for Improving and Predicting Yield in Wafer Fabrication System," Journal of Korea Intelligent Information Systems Society, Vol. 9, No. 1, pp. 157-177, 2003.

[3] Byeon, S. K., Kang, C. W., and Sim, S. B., "Defect Type Prediction Method in Manufacturing Process Using Data Mining Technique," Journal of the Society of Korea Industrial and Systems Engineering, Vol. 27, No. 2, pp. 10-16, 2004.

[4] Han, J., Pei, J., and Yin, Y., "Mining Frequent Patterns Without Candidate Generation," ACM Sigmod Record, Vol. 29, No. 2, pp. 1-12, 2000.

[5] Jacobs, A., "The Pathologies of Big Data," Communications of the Association for Computing Machinery, Vol. 52, No. 8, pp. 36-44, 2009.

[6] Jin, H. J., Kim, H. T., and Lee, Y. H., "A Context-Aware System for Reliable RFID-Based Logistics Management," Journal of Society for e-Business Studies, Vol. 18, No. 2, pp. 223-240, 2013.

[7] Kang, S. K. and Kim, B. T., Preface, The Journal of The Korean Institute of Communication Sciences, Vol. 35, No. 1, pp. 4-4, 2017.

[8] Kim, D. S., Wong, W. X. S., Lim, M. S., Liu, C., Kim, N. G., Park, J. Y., Kil, W. Y., and Yoon, H. S., "A Methodology for Analyzing Public Opinion about Science and Technology Issues Using Text Analysis," Journal of Information Technology Services, Vol. 14, No. 3, pp. 33-48, 2015.

[9] Kim, D. S. and Kim, N. G., "Mapping Categories of Heterogeneous Sources Using Text Analytics," Journal of Intelligent Information Systems, Vol. 22, No. 4, pp. 193-215, 2016.

[10] Kim, K. Y. and Bae, S. J., "Establishing Method of RAM Objective Considering Combat Readiness and Field Data of Similarity Equipment," Journal of the Society of Korea Industrial and Systems Engineering, Vol. 32, No. 3, pp. 127-134, 2009.

[11] Kim, N. G., Lee, D. H., Choi, H. C., and Wong, W. X. S., "Investigations on Techniques and Applications of Text Analytics," The Journal of Korean Institute of Communications and Information Sciences, Vol. 42, No. 2, pp. 471-492, 2017.

[12] Kim, S. R., Jang, G. S., and Cho, C. W., "Case Study of Design and Implementation for Hadoop-Based Integrated Facility Monitoring System," Journal of Korean Institute of Industrial Engineers, Vol. 40, No. 1, pp. 34-42, 2014.

[13] Kim, S. Y. and Song, M. K., "Application of MI-NPS Digital Factory Methodology for Production Ability Improvement and Optimal Layout Design: Applied Case to Vehicle Shaft Manufacturing Line," Korea

Productivity Association, Vol. 28, No. 1, pp. 47-73, 2014.

[14] KTPM Management Consulting, Failure Zero Propulsion Manual, 2001.

[15] Douglas, L., 3D Data Management: Controlling Data Volume, Velocity, and Variety, META Group, 2001.

[16] Lee, I. M., Yoo, S. B., Joung, H. S., and No, S. H., "Study on Power Equipment Maintenance and Replacement Cycle," In Proceedings of the Korean Institute of Electrical Engineers, pp. 1460-1460, 2015.

[17] Lee, D. H., Park, J. H., and Bae, H. R., "Comparison Between Planned and Actual Data of Block Assembly Process Using Process Mining in Shipyards," Journal of Society for e-Business Studies, Vol. 18, No. 4, pp. 145-167, 2013.

[18] Lee, S. W. and Lee, J. K., "Equipment Management Through Equipment Direct Interface," In Proceedings of Academic Conference Symposium, pp. 523-524, 2015.

[19] Lee, W. B. and Kim, S. K., "An Empirical Research on the Impact of Growth in a National Industrial Complex on Job Creation," Journal of The Korean Regional Development Association, Vol. 26, No. 2, pp. 79-100, 2014.

[20] Lim, M. S. and Kim, N. G., "Investigating Dynamic Mutation Process of Issues Using Unstructured Text Analysis," Journal of Intelligent Information Systems, Vol. 22, No. 1, pp. 1-18, 2016.

[21] Lim, J. W., Jo, D. H., Lee, S. Y., Park, H. J., and Park, J. W., "A Case Study for the Smart Factory Application in the Manufacturing Industry," Korean Journal of Business Administration, Vol. 30, No. 9, pp. 1609-1630, 2017.

[22] Noh, K. S. and Park, S. H., "An Exploratory Study on Application Plan of Big Data to Manufacturing Execution System," Journal of Digital Convergence, Vol. 12, No. 1, pp. 305-311, 2014.

[23] Pak, H. W., "Trends in Production and Manufacturing Technologies Related to Smart Factories," Korea Institute of Communication Sciences, Vol. 33, No. 1, pp. 24-29, 2015.

[24] Park, C. G., Roh, H. S., Choi, Y. J., Kim, H. W., and Lee, J. K., "A Study on the Application Methods of Big Data in the Technology Commercialization Process," Journal of Society for e-Business Studies, Vol. 19, No. 4, pp. 73-99, 2014.

[25] Park, E. J. and Byung-Hyun, H. A., "A Formal Framework for Analyzing Performance of Container Terminal Operations," Journal of Society for e-Business Studies, Vol. 18, No. 2, pp. 191-203, 2013.

[26] Park, J. B. and Bae, Y. J., "A Study on Management System for Improving of Equipment Productivity in Injection Machining Process," Journal of the Korea Safety Management & Science, Vol. 19, No. 1, pp. 219-226, 2017.

[27] Schwab, K., The Fourth Industrial Revolution, Crown Business, 2017.

[28] Seo, M. K. and Yun, W. Y., “Clustering-Based Monitoring and Fault Detection in Hot Strip Roughing Mill,” Journal of the Korean Society for Quality Management, Vol. 45, No. 1, pp. 25-38, 2017.

[29] Taner, T. and Sivrioglu, M., “Energy-Exergy Analysis and Optimization of a Model Sugar Factory in Turkey,” Energy, Vol. 93, No. 1, pp. 641-654, 2015.

[30] Kabitzsch, K., Vasyutynskyy, V., and Kotte, G., “Troubleshooting in MES of Semiconductor Factory,” In Proceedings of the International Conference on Production Research, 2005.

[31] Yun, Y. G. and Gang, Y. S., “A Study on Estimation Flow of Information Analysis for Prevention of Human Error to the Operation,” In Proceedings of the Safety Management and Science Conference, pp. 231-241, 2013.

저 자 소 개

윤준서 (E-mail: jsyun2009@gmail.com)
2018년 경기대학교 산업경영공학과 (학사)
2018년~현재 핑거포인트랩 대리
관심분야 자연어 처리, 빅데이터 기반의 인간 모델링

안현태 (E-mail: hyeontaean94@gmail.com)
2013년~현재 경기대학교 산업경영공학과 (학사)
관심분야 텍스트 마이닝

최예림 (E-mail: yrchoi@kgu.ac.kr)
2010년 서울대학교 산업공학과 (학사)
2016년 서울대학교 산업공학과 (박사)
2016년~2017년 네이버랩스 Data Scientist
2017년~현재 경기대학교 산업경영공학과 조교수
관심분야 인공지능/머신러닝, 빅데이터 기반의 인간 모델링