



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2022년02월09일  
(11) 등록번호 10-2360362  
(24) 등록일자 2022년02월04일

(51) 국제특허분류(Int. Cl.)  
G05B 23/02 (2006.01) G05B 19/418 (2006.01)  
G06N 3/08 (2006.01)  
(52) CPC특허분류  
G05B 23/024 (2013.01)  
G05B 19/41875 (2013.01)  
(21) 출원번호 10-2020-0046721  
(22) 출원일자 2020년04월17일  
심사청구일자 2020년04월17일  
(65) 공개번호 10-2021-0128713  
(43) 공개일자 2021년10월27일  
(56) 선행기술조사문헌  
KR1020170024262 A\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
충북대학교 산학협력단  
충청북도 청주시 서원구 충대로 1 (개신동)  
(72) 발명자  
팽틀라  
충청북도 청주시 서원구 충대로 1 충북대학교  
류관희  
대전 유성구 엑스포로 448 302-1602  
(74) 대리인  
김정현

전체 청구항 수 : 총 5 항

심사관 : 김윤환

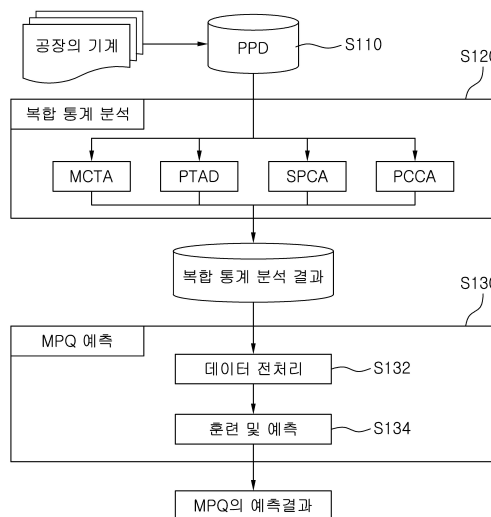
(54) 발명의 명칭 통계적 분석과 딥러닝을 이용한 제조 공정 품질 분석 및 예측 방법 및 이를 기록한 기록매체

(57) 요약

본 발명은 제조 공정 품질 분석 및 예측 방법에 관한 것으로서, 공장의 기계로부터 생산되는 제품 공정에 관련된 제품 공정 데이터를 실시간으로 수집하는 단계, 수집된 제품 공정 데이터에 대해 통계적 분석을 수행하여 제품 제조 공정을 분석하고, 이를 통해 비정상 공정을 검출하는 통계적 분석 단계 및 통계적 분석 결과 데이터를 이용하여 제조 공정 품질을 예측하는 예측 단계를 포함한다.

본 발명에 의하면, 통계적 분석과 딥러닝을 이용하여 제조 공정 품질을 분석하고 예측함으로써, 비정상적인 공정을 탐색하고, 제조 공정 품질을 예측하여 제품 생산 효율을 향상시킬 수 있는 효과가 있다.

대표도 - 도1



(52) CPC특허분류  
*G05B 23/0283* (2013.01)  
*G06N 3/08* (2013.01)

이 발명을 지원한 국가연구개발사업

|             |                                    |
|-------------|------------------------------------|
| 과제고유번호      | 1415167171                         |
| 과제번호        | 20004367                           |
| 부처명         | 산업통상자원부                            |
| 과제관리(전문)기관명 | 한국산업기술평가원                          |
| 연구사업명       | 산업핵심기술개발사업                         |
| 연구과제명       | 세라믹산업 제조혁신을 위한 클라우드 기반 빅데이터 플랫폼 개발 |
| 기여율         | 1/1                                |
| 과제수행기관명     | 충북대학교                              |
| 연구기간        | 2020.01.01 ~ 2020.12.31            |

---

## 명세서

### 청구범위

#### 청구항 1

공장의 기계로부터 생산되는 제품 공정에 관련된 제품 공정 데이터를 실시간으로 수집하는 단계;  
 수집된 제품 공정 데이터에 대해 통계적 분석을 수행하여 제품 제조 공정을 분석하고, 이를 통해 비정상 공정을 검출하는 통계적 분석 단계; 및  
 통계적 분석 결과 데이터를 이용하여 제조 공정 품질을 예측하는 예측 단계를 포함하고,  
 상기 통계적 분석 단계에서,  
 제품 제조 주기 시간을 분석하는 MCTA(manufacturing cycle time analysis), 공정 궤적의 비정상을 탐지하는 PTAD(process trajectory abnormal detection), 통계적 공정 제어를 분석하는 SPCA(Statistical Process Control Analysis) 및 공정 능력 제어를 분석하는 PCCA(Process Capability Control Analysis)로 구성된 복합 통계적 분석을 수행하여 제품 제조 공정을 분석하는 것을 특징으로 하는 제조 공정 품질 분석 및 예측 방법.

#### 청구항 2

청구항 1에 있어서,  
 상기 제품 공정 데이터는 기계 작동 상태, 생산 공정, 기계의 조건 값 및 기계의 테스트 결과 데이터를 포함하는 것임을 특징으로 하는 제조 공정 품질 분석 및 예측 방법.

#### 청구항 3

삭제

#### 청구항 4

청구항 1에 있어서,  
 상기 예측 단계에서,  
 통계적 분석 결과 데이터에 대하여 숫자 값이 아닌 레이블 값을 포함하는 변수인 범주형 데이터로 인코딩하는 범주형 데이터 인코딩 과정과,  
 트레이닝 세트, 유효성 검사 세트 및 테스트 세트로 구분하는 데이터 세트 분할 과정과,  
 데이터 기능의 범위를 정규화하는 기능 스케일링 과정을 포함하는 데이터 전처리 단계를 수행하는 것을 특징으로 하는 제조 공정 품질 분석 및 예측 방법.

#### 청구항 5

청구항 4에 있어서,  
 상기 예측 단계에서,  
 상기 데이터 전처리 단계를 수행한 후,  
 LSTM(Long Short-Term Memory) 모델을 기반으로 전처리된 데이터를 트레이닝하고, 이를 통해 제조 공정 품질을 예측하는 단계를 수행하는 것을 특징으로 하는 제조 공정 품질 분석 및 예측 방법.

**청구항 6**

청구항 1, 청구항 2, 청구항 4, 청구항 5 중 어느 한 청구항의 방법을 컴퓨터로 실행시킬 수 있는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

**발명의 설명**

**기술분야**

[0001] 본 발명은 제조 공정에서의 품질 분석 및 예측 기술에 관한 것이다.

**배경기술**

[0002] 4차 산업 혁명 (Industry 4.0)은 현재 제조 산업의 자동화 및 데이터 교환 추세를 나타내고, 제조 관리에서 개발 공정 및 생산 체인을 의미한다. 최근에는 제조 공정이 점점 더 복잡해지고 있다. 제조 공정이 복잡해 지고 지속적으로 증가하는 데이터 양으로 인해 생산 공정 모니터링 및 데이터 분석과 관련하여 제조업체에 과도한 요구가 발생한다. 그리고, 제조 공정에서 품질 관리는 항상 필수 요소이다.

[0003] 사물 인터넷(IoT)을 사용하면 네트워크 센서 및 지능형 장치를 사용하여 제조 공정의 모든 지점에서 데이터를 수집하고, 이러한 기술을 제조 현장에서 직접 사용할 수 있으며, 인공 지능(artificial intelligence, AI) 및 예측 분석을 유도하기 위해 데이터를 수집 할 수 있다.

[0004] 이처럼, 데이터 양이 크게 증가하면서 제품 품질 관리 및 예측 유지 보수 (predictive maintenance, PdM)에 새로운 패러다임이 만들어졌다. PdM은 Industry 4.0의 핵심 이점 중 하나인데, 잠재적인 문제를 분석, 탐지 및 예측하기 위해 각 시스템에서 가져온 특정 정보에 의존하며, 생산성을 높이고 제품 품질을 향상 시키며 제조업체에 신뢰성을 제공한다.

[0005] 예측 가능한 유지 보수, 생산 모니터링 및 품질 관리는 서로 밀접하게 연결되어 있다. 생산 모니터링 및 품질 관리는 모든 제조 유형에 있어 중요한 비즈니스 전략이 되었다. 최근 수십 년 동안 품질을 제어하고 제조 공정을 가속화하기 위한 많은 방법이 개발되었다. 품질 관리는 개념에서부터 최종 제품의 납품에 이르기까지 제조 공정의 모든 부분에 필수적이다. 이와 유사하게, PdM에서 장애를 예측하기 위해 패턴을 정의하는 것을 목표로, 특정 딥 러닝 알고리즘으로부터 이익을 얻을 수 있는 공정인 제품 생산 상태를 모니터링하기 위해, 시간이 지남에 따라 데이터가 수집된다.

[0006] 딥 러닝(deep learning)은 방대한 양의 데이터로 학습 할 때 뛰어난 정확도로 인해 많은 인기를 얻고 있다. 1997년에 LSTM(Long Short-Term Memory)이라는 재귀 신경망(recurrent neural network, RNN)의 개선된 버전이 소실 구배 문제를 해결하고 복잡한 시간 시퀀스 데이터를 처리하기 위해 제안되었다. LSTM은 많은 종류의 작업에 있어 전통적인 RNN보다 긴 일련의 데이터가 포함된 작업에 훨씬 더 효과적인 특수 종류의 인공 RNN이다. 딥 러닝 알고리즘 중에서 LSTM 네트워크는 복잡한 시퀀스를 학습하는 데 능숙하고 고장 패턴을 감지하는 데 오랜 시간 동안 기능하기 때문에 PdM에 특히 매력적이다.

[0007] 요즘 PdM은 스마트 제조 분야에서 중요한 연구 분야가 되고 있다. 많은 PdM 시스템이 제조 산업에 적용되도록 제안되고 개발되었다. 대부분의 이전 연구는 머신에서 수집한 이력 오류 데이터를 사용하여 머신의 상태에 중점을 두었다.

[0008] 이처럼, 제조 산업에서 생산 공정 관리는 제조 공정 성능의 품질을 지속적으로 모니터링하여 고품질 제품을 생산하는 동안 생산 공정 안정성을 보장하는 데 매우 중요하다. 실제로, 모든 제조 산업에서 만족스러운 수익을 창출하려면 고품질 생산이 필요하다.

**선행기술문헌**

**특허문헌**

[0009] (특허문헌 0001) 대한민국 등록특허 10-1661818

**발명의 내용**

**해결하려는 과제**

- [0010] 본 발명은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 본 발명에서는 상태 모니터링을 통한 제품 생산 공정에 중점을 두고, 기계로부터 제품 작동 데이터를 수집하고, 제품 고장으로 이어질 수 있는 공정 조건을 분석하고, 변경을 감지하고, 제조 공정 품질(MPQ)의 향후 안정성을 예측하는데 그 목적이 있다.
- [0011] 본 발명의 목적은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

**과제의 해결 수단**

- [0012] 이와 같은 목적을 달성하기 위한 본 발명은 제조 공정 품질 분석 및 예측 방법에 관한 것으로서, 공장의 기계로부터 생산되는 제품 공정에 관련된 제품 공정 데이터를 실시간으로 수집하는 단계, 수집된 제품 공정 데이터에 대해 통계적 분석을 수행하여 제품 제조 공정을 분석하고, 이를 통해 비정상 공정을 검출하는 통계적 분석 단계 및 통계적 분석 결과 데이터를 이용하여 제조 공정 품질을 예측하는 예측 단계를 포함한다.
- [0013] 상기 제품 공정 데이터는 기계 작동 상태, 생산 공정, 기계의 조건 값 및 기계의 테스트 결과 데이터를 포함하는 것일 수 있다.
- [0014] 상기 통계적 분석 단계에서, 제품 제조 주기 시간을 분석하는 MCTA(manufacturing cycle time analysis), 공정 궤적의 비정상 탐지하는 PTAD(process trajectory abnormal detection), 통계적 공정 제어를 분석하는 SPCA(Statistical Process Control Analysis) 및 공정 능력 제어를 분석하는 PCCA(Process Capability Control Analysis)로 구성된 복합 통계적 분석을 수행하여 제품 제조 공정을 분석할 수 있다.
- [0015] 상기 예측 단계에서, 통계적 분석 결과 데이터에 대하여 숫자 값이 아닌 레이블 값을 포함하는 변수인 범주형 데이터로 인코딩하는 범주형 데이터 인코딩 과정과, 트레이닝 세트, 유효성 검사 세트 및 테스트 세트로 구분하는 데이터 세트 분할 과정과, 데이터 기능의 범위를 정규화하는 기능 스케일링 과정을 포함하는 데이터 전처리 단계를 수행할 수 있다.
- [0016] 상기 예측 단계에서, 상기 데이터 전처리 단계를 수행한 후, LSTM(Long Short-Term Memory) 모델을 기반으로 전처리된 데이터를 트레이닝하고, 이를 통해 제조 공정 품질을 예측하는 단계를 수행할 수 있다.

**발명의 효과**

- [0017] 본 발명에 의하면, 통계적 분석과 딥러닝을 이용하여 제조 공정 품질을 분석하고 예측함으로써, 비정상적인 공정을 탐색하고, 제조 공정 품질을 예측하여 제품 생산 효율을 향상시킬 수 있는 효과가 있다.

**도면의 간단한 설명**

- [0018] 도 1은 본 발명의 일 실시예에 따른 제조 공정 품질 예측을 위한 시스템 아키텍처를 도시한 것이다.
- 도 2는 본 발명의 일 실시예에 따른 제조 공정 품질을 추정하는 데 사용되는 PPD를 예시한 도표이다.
- 도 3은 제조 주기 시간을 설명하기 위한 도면이다.
- 도 4는 하나의 기계에서 다수의 공정 궤적을 도시한 것이다.
- 도 5는 공정 능력 관리를 분석한 그래프이다.
- 도 6은 공정 능력(Cp)의 기준을 보여주는 도표이다.
- 도 7은 공정 능력 지수(Cpk)의 기준을 보여주는 도표이다.
- 도 8은 본 발명의 일 실시예에 따른 공정 품질 성능을 결정하기 위해 LSTM을 적용하는 절차를 도시한 것이다.
- 도 9는 본 발명의 일 실시예에 따른 PPD 데이터에 복합 통계적 분석을 적용한 결과를 보여주는 도표이다.

도 10은 MPQ의 기준을 예시한 도표이다.

도 11은 혼동 매트릭스의 네가지 주요 요소를 보여주는 도표이다.

도 12는 혼동 매트릭스의 오류 매트릭을 나타낸 도표이다.

**발명을 실시하기 위한 구체적인 내용**

- [0019] 본 명세서에서 개시된 실시예의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 개시에서 제안하고자 하는 실시예는 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 당해 기술분야에서 통상의 지식을 가진 자에게 실시예들의 범주를 완전하게 알려주기 위해 제공되는 것일 뿐이다.
- [0020] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 개시된 실시예에 대해 구체적으로 설명하기로 한다.
- [0021] 본 명세서에서 사용되는 용어는 개시된 실시예들의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 관련 분야에 종사하는 기술자의 의도 또는 관례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 명세서의 상세한 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 명세서의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [0022] 본 명세서에서의 단수의 표현은 문맥상 명백하게 단수인 것으로 특정하지 않는 한, 복수의 표현을 포함한다.
- [0023] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 명세서에서 사용되는 "부"라는 용어는 소프트웨어, FPGA 또는 ASIC과 같은 하드웨어 구성요소를 의미하며, "부"는 어떤 역할들을 수행한다. 그렇지만 "부"는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. "부"는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 "부"는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 공정들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 "부"들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 "부"들로 결합되거나 추가적인 구성요소들과 "부"들로 더 분리될 수 있다.
- [0024] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0025] 본 발명에서는 상태 모니터링을 통한 제품 생산 공정에 중점을 두고, 기계로부터 제품 작동 데이터를 수집하고, 제품 고장으로 이어질 수 있는 공정 조건을 분석하고, 변경을 감지하고, 제조 공정 품질(manufacturing process quality, MPQ)의 향후 안정성을 예측한다.
- [0026] MPQ는 비정상 공정을 결정하고 향후 제조 공정의 품질을 예측하기 위해 제품 생산 공정을 분석하는 데 사용되는 방법으로서, 자동차 제조 산업의 기계에서 생산 공정 운영에 대한 수집된 데이터를 사용하여 비정상 공정 품질을 결정할 수 있다.
- [0027] 본 발명에서는 MCTA(manufacturing cycle time analysis), PTAD(process trajectory abnormal detection), SPCA(statistical process control analysis) 및 PCCA(process capability control analysis)와 같은 복합 통계 공정 분석을 사용하여 기계에서 제품을 생산하는 공정을 분석하고, 식별 불량품으로 이어질 수 있는 품질이 낮은 비정상 공정을 검출한다.
- [0028] 그리고, 본 발명에서는 복합 통계 공정 분석 결과 및 LSTM 신경망을 사용하여 향후 제조 공정 품질을 예측한다.
- [0029] 본 발명은 제조 공정 품질 분석 및 예측 방법에 대한 것이다.
- [0030] 본 발명의 제조 공정 품질 분석 및 예측 방법을 수행하는 주체는 제조 공정 품질 분석 및 예측 방법을 수행하는 제반 컴퓨터 장치 또는 서버 시스템이라고 할 수 있으며, 또는 제조 공정 품질 분석 및 예측 방법을 수행하는 시스템 또는 장치를 전반적으로 제어하는 제어부나 프로세서(processor)일 수 있다. 즉, 본 발명의 제조 공정 품질 분석 및 예측 방법은 일종의 소프트웨어인 알고리즘으로 구성될 수 있으며, 소프트웨어는 제조 공정 품질

분석 및 예측 방법을 수행하는 서버 시스템, 장치의 제어부 또는 프로세서(processor)에서 실행될 수 있다.

- [0031] 이제 본 발명에서 제조 공정 품질(MPQ)을 분석하고 예측하기 위한 방법에 대한 시스템 아키텍처에 대해 설명한다.
- [0032] 도 1은 본 발명의 일 실시예에 따른 제조 공정 품질 예측을 위한 시스템 아키텍처를 도시한 것이다.
- [0033] 도 1을 참조하면, 본 발명의 시스템은 실시간 데이터 수집 단계(S110), 복합 통계적 분석 단계(S120), LSTM을 이용한 MPQ 예측 단계(S130)의 세가지 구성요소로 구분할 수 있다.
- [0034] 데이터 수집 단계(S110)는 공장의 기계로부터 생산되는 제품 공정에 관련된 제품 공정 데이터(Production Process Data, PPD)를 실시간으로 수집하는 단계이다.
- [0035] 통계적 분석 단계(S120)는 수집된 제품 공정 데이터에 대해 통계적 분석을 수행하여 제품 제조 공정을 분석하고, 이를 통해 비정상 공정을 검출하는 단계이다.
- [0036] 예측 단계(S130)는 통계적 분석 결과 데이터를 이용하여 제조 공정 품질을 예측하는 단계이다.
- [0037] 본 발명의 일 실시예에서 제품 공정 데이터(PPD)는 기계 작동 상태, 생산 공정, 기계의 조건 값 및 기계의 테스트 결과 데이터를 포함하는 것일 수 있다.
- [0038] 본 발명에서 통계적 분석 단계(S120)에서, 제품 제조 주기 시간을 분석하는 MCTA(manufacturing cycle time analysis), 공정 궤적의 비정상 탐지하는 PTAD(process trajectory abnormal detection), 통계적 공정 제어를 분석하는 SPCA(Statistical Process Control Analysis) 및 공정 능력 제어를 분석하는 PCCA(Process Capability Control Analysis)로 구성된 복합 통계적 분석을 수행하여 제품 제조 공정을 분석할 수 있다.
- [0039] 예측 단계(S130)는 데이터 전처리 단계(S132)와, 트레이닝 및 예측 단계(S134)로 구성된다.
- [0040] 전처리 단계(S132)는 통계적 분석 결과 데이터에 대하여 숫자 값이 아닌 레이블 값을 포함하는 변수인 범주형 데이터로 인코딩하는 범주형 데이터 인코딩(Encoding categorical data) 과정과, 트레이닝 세트, 유효성 검사 세트 및 테스트 세트로 구분하는 데이터 세트 분할(Splitting dataset) 과정과, 데이터 기능의 범위를 정규화하는 기능 스케일링(Feature scaling) 과정을 포함하여 이루어진다.
- [0041] 그리고, 트레이닝 및 예측 단계(S134)는 데이터 전처리 단계(S132)를 수행한 후, LSTM(Long Short-Term Memory) 모델을 기반으로 전처리된 데이터를 트레이닝하고, 이를 통해 제조 공정 품질을 예측한다.
- [0042] 일반적으로 제품을 제조하는 동안 공장의 기계는 기계 작동 상태, 생산 공정, 기계의 조건 값 및 테스트 결과를 포함하는 다양한 유형의 데이터를 생성한다. 본 발명에서 이러한 데이터는 PPD(Production Process Data)로 표현된다.
- [0043] PPD에는 기계에 대한 규격 관리 하한 및 상한뿐만 아니라 기계에서 생산하기 위한 수량, 조건 값, 생성 시작 및 종료 시간이 포함될 수 있다. PPD 세트는 제품을 생산하는 동안 제조 산업의 기계에서 수집된다. 일반적으로 기계에서 수집한 데이터는 무선 또는 유선 네트워크를 통해 서버 시스템으로 실시간 전송된다. 이를 위해 Apache Kafka 스트리밍 개선 시스템을 사용할 수 있다. 이 시스템은 일반적으로 데이터를 스트리밍하고 전송하는 방법과 실시간 데이터 스트리밍의 목표를 달성하기 위해 빅 데이터 생태계와 데이터 전송 보안을 활용하는 전이 패턴을 사용한다.
- [0044] 그리고, 주기적으로 제품을 생산하는 각 기계에서의 비정상 공정을 탐지하기 위해 PPD의 복합 통계적 분석(composite statistical analysis)이 제안된다.
- [0045] 보다 나은 분석을 제공하기 위해 MCTA(manufacturing cycle time analysis), PTAD(process trajectory abnormal detection), SPCA(Statistical Process Control Analysis) 및 PCCA(Process Capability Control Analysis)를 비롯한 4 가지 주요 구성요소를 포함한다.
- [0046] MCTA는 제품을 만드는 데 걸리는 시간으로 정의된다.
- [0047] PTAD는 기계가 제품을 생산할 때 비정상적인 공정을 식별한다.
- [0048] SPCA에서 주요 부분은 공정 품질을 제어하는 데 사용되는 X 및 R 유형의 관리도이다.
- [0049] 마지막으로, PCCA의 Cp와 Cpk는 공정의 품질을 측정하고 공정이 얼마나 잘 실행되고 있는지 측정하기 위해 적용된다.

- [0050] 셋째, PPD의 복합 통계적 분석에서 추출된 원본 데이터 세트의 일부 유용한 데이터와 기능은 LSTM 예측 모델의 입력으로 사용되어 다음 공정 순서를 예측한다.
- [0051] 그 다음, 기능 선택 및 기능 추출을 사용하여 데이터 전처리 단계를 수행한다. 여기서 범주형 데이터를 인코딩하고 데이터 집합과 함께 사용하여 기능을 확장할 수 있다.
- [0052] 그 다음, 모든 기능이 LSTM 알고리즘에 의한 트레이닝 및 예측에 사용된다.
- [0053] 마지막으로, 소프트맥스(Softmax) 활성화 기능은 예측 모델의 마지막 레이어와 함께 작동하여 확률 결과를 0.00에서 1.00으로 분류하고 출력한다. 결과가 높을수록 우수한 공정 품질을 나타내며, 불량품 출력이 없거나 적을 수 있다.
- [0055] 이제 본 발명에서 복합 통계적 분석에 대해 설명하고자 한다.
- [0056] 복합 통계적 분석에서는 PPD(Production Process Data) 세트를 이용하여 생산 공정의 품질을 분석한다. 특히, 그 결과는 기계에서 비정상적인 공정을 감지하는데 사용된다. 분석은 제조주기 시간 분석(manufacturing cycle time analysis, MCTA), 공정 궤적 비정상 탐지(process trajectory abnormal detection, PTAD), 통계적 공정 제어 분석(Statistical Process Control Analysis, SPCA) 및 공정 능력 제어 분석(Process Capability Control Analysis, PCCA)을 포함한 4 가지 분석 항목으로 구성된다.
- [0057] 일반적으로 주어진 작업 시간 동안 동일한 종류의 제품을 정기적으로 제조하는 공정을 고려하면, 이제 제조 공정의 MPQ(Process Quality Performance)를 알고자 한다. MPQ를 보다 정확하게 분석하고 예측하려면 데이터가 필요하다. 이를 위하여 기계에서 생성된 데이터를 사용하며, 이 데이터를 생산 공정 데이터(PPD)라고 한다.
- [0058] 우선, 작업 시작 및 종료 시간이 PPD에 포함된다. 그리고 주어진 작업 시간 동안 정상 및 비정상 제품의 수가 PPD에 포함된다.
- [0059] 압력, 온도 및 균형과 같은 여러 조건 변수는 기계로 만든 단일 제품과 관련이 있다. 조건 변수는 공작물 시작 및 종료 시간, 공정 조건 변수 명칭, 해당 명칭의 공정 시작 및 종료 시간, 해당 명칭의 실제 데이터, 해당 명칭의 규격 상한(upper specification limit, USL) 및 규격 하한(lower specification limit, LSL), 및 공정 결과를 포함한다.
- [0060] PPD 세트는 제품을 생산하는 동안 제조 공정의 기계로부터 수집된다. 본 발명에 사용된 PPD의 세부 사항은 도 2에 기술되어 있다.
- [0061] 도 2는 본 발명의 일 실시예에 따른 제조 공정 품질을 추정하는 데 사용되는 PPD를 예시한 도표이다.
- [0062] 제조 주기 시간 분석(MCTA)에 대해 설명하면 다음과 같다.
- [0063] 일반적으로 제조 주기 시간(manufacturing cycle time, MCT)은 제조 산업에서 중요한 성능 측정 수단으로 널리 알려져 있다. MCT는 관련된 생산 공정 단계 수 때문에 복잡해졌으며, 각 부품의 지속 시간을 파악하기 위해 제품 주기 시간을 공작물로 세분화함으로써 시스템 개선이 가능해졌다. 이는 작동 중인 공작물이 제품 주기 시간을 개선하고, 각 제품의 생산 기간에 대한 이해를 높일 수 있다는 것을 의미한다. 이를 통해 우수 및 결함 제품 생산에 있어 지속시간 패턴을 분석할 수 있는 제조주기 시간 분석이 제안된다. 여기에는 도 3과 같이 가공품 주기 시간, 순수 주기 시간, 공정 주기 시간 등 4가지 요소가 포함된다.
- [0064] 도 3은 제조 주기 시간을 설명하기 위한 도면이다.
- [0065] 이 단계에서는 어떤 제품의 반복을 수행하는 데 걸리는 시간을 분석하는데, 일반적으로 기계에 있는 한 제품의 시작점에서 같은 기계에 있는 다른 제품의 시작점까지 측정한다.
- [0066] 공작물 사이클 시간(Workpiece cycle time, WCt)은 단일 제품을 만들기 위해 제품 구성 요소를 기계에 로드(load)하는 시간이다. 이 로딩 공정은 이전 WCt 종료 시간 전에 시작되거나 또는 이후에 시작될 수 있다. 공작물 사이클 시간의 안정성을 나타내는 sWCt의 계산은 다음 공식과 같다.

[0067] 
$$sWCt = 1/n \sum_{i=0}^n |(wet_i - wst_i) - MWct| \quad (1)$$

[0068] 여기서, n은 총 제품 수이고, wst<sub>i</sub>는 i번째 제품에 대한 공작물 사이클의 시작 시간이고, wet<sub>i</sub>는 i번째 제품에

대한 공작물 사이클의 종료 시간이고, MWCt는 모든 n 공작물 사이클 시간의 평균이다.

[0069] 순수 사이클 시간(pure cycle time, PrCt)은 조건 변수가 있는 단일 제품을 순수하게 생산하는 시간이다. 하나의 PrCt에는 많은 조건 변수의 값이 있다. 순수한 사이클 시간의 안정성을 나타내는 sPrCt의 계산은 다음 공식과 같이 나타낸다.

$$sPrCt = 1/n \sum_{i=0}^n |(pet_i - pst_i) - MPrCt| \quad (2)$$

[0071] 여기서 n은 총 제품 수이고, pst<sub>i</sub> 및 pet<sub>i</sub>는 각각 i 번째 제품에 대한 제품주기의 시작 시간 및 종료 시간이고, MPrCt는 모든 n 순수 사이클 시간의 평균이다.

[0072] 순수 간격 시간(Pure interval time, PrIt)은 제품 완료와 다음 제품 시작 사이의 시간을 나타낸다. 순수 간격 시간의 안정성 sPrIt은 다음 식으로 계산된다.

$$sPrIt = 1/n \sum_{i=0}^{n-1} |(pst_{i+1} - pet_i) - MPrIt| \quad (3)$$

[0074] 여기서 n은 총 제품 수이고, pst<sub>i+1</sub>는 i+1 번째 제품의 시작 시간이고, pet<sub>i</sub>는 i 번째 제품의 종료 시간이고, MPrCt는 모든 n 순수 간격 시간의 평균이다.

[0075] 공정 사이클 시간(Process cycle time, PCt)은 기계에서 한 공작물 사이클 시간의 공정 시작부터 동일한 기계에서 다른 공작물 사이클 시간의 시작까지의 지속 시간이다. 공정 사이클 시간의 안정성 sPCt는 다음 식에 의해 계산된다.

$$sPCt = 1/n \sum_{i=0}^{n-1} |(pst_{i+1} - pst_i) - MPCt| \quad (4)$$

[0077] 여기서, n은 총 제품 수이며, pst<sub>i+1</sub>는 i+1 제품의 시작 시간이고, pet<sub>i</sub>는 i 번째 제품의 종료 시간이고, MPCt는 모든 n 공정 주기 시간의 평균이다.

[0078] 이제, PTAD(Process Trajectory Abnormal Detection)에 대해 설명하고자 한다.

[0079] 궤적 비정상 탐지(Trajectory abnormal detection, TAD)는 서로 다른 시나리오에서 유용한 분석 방법이며, 비정상 궤적, 비정상 서브 궤적, 비정상 도로 구간, 비정상 이벤트 및 시계열 데이터의 비정상 이동 물체와 같은 영역을 고려한다. 기존에 많은 연구들을 통해 방향, 거리 기반 접근법, 역사적 유사성 기반 접근법, 밀도 기반 및 분류 기반 접근법을 포함하는 TAD 기법을 제안한 바 있다. 본 발명에서는 가우스 정규 분포법을 이용하여 비정상적인 공정의 검출에 생산 공정 조건 데이터를 적용한다.

[0080] PTAD(Process Trajectory Abnormal Detection) 분석을 통해 기계에서 생산 공정 상태 이동의 동작을 이해할 수 있다. 도 4는 많은 공정 조건을 갖는 기계를 보여주고 있다.

[0081] 도 4는 하나의 기계에서 다수의 공정 궤적을 도시한 것이다.

[0082] 도 4에서 각 공정 조건은 t<sub>1</sub>, t<sub>2</sub>, t<sub>3</sub>, ..., t<sub>n</sub>으로 표시되는 궤적을 고려한다.

[0083] 다음은 가우스 분포 방정식이다.

$$GD_g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

[0085] 여기서 x는 연속 랜덤 변수, e는 지수, μ는 랜덤 변수 x의 평균, σ는 랜덤 변수 x의 표준 편차이다. 매개변수 μ 및 σ는 가우스 곡선의 중심 및 확산을 각각 나타낸다. 또한 밀도는 평균 주위에서 더 높고, 평균으로부터 거리가 멀어 질수록 빠르게 감소한다는 것을 알 수 있다. 이 모델 포인트는 평균값을 중심으로 분산되고, 분산으로 확산되므로 가우스를 정의하는데 평균과 분산의 두 값이 필요하다.

[0086] 공정 궤적이 비정상인지 여부를 확인하기 위해, 우선 정상 제품의 실제 데이터를 사용하여 정규 가우스 분포

$GD_{normal}(t)$ 를 계산한다.

[0087] 다음으로, 지속 시간을 나타내는 그룹에 대한 PTAD를 측정하기 위해, 그룹 동안 생산된 모든 제품과 관련된 실제 데이터를 갖는 가우스 분포  $GD_{group}(t)$ 를 구한다.  $i$ 번째 그룹의  $PTAD_i$ 는 다음과 같이 계산된다.

$$PTAD_i = 1/t \sum_{k=1}^t |GD_i(k) - GD_{normal}(k)| \quad (6)$$

[0088] 따라서 가우시안 하에서 새로운 데이터의 확률이 임계 값보다 높거나 낮 으면, 임계 값을 벗어난 비정상 공정일 수 있다.

[0089] 통계적 공정 관리(Statistical process control, SPC)는 제조 생산 라인에서 공정 품질을 측정하고 제어하는 통계 기법으로서, 기계가 제품을 만드는 동안 공정을 유지하기 위해 제조 산업에서 광범위하게 사용된다.

[0092] 관리도(control chart)는 SPC에서 가장 중요한 도구 중 하나이며, 생산 공정의 품질을 관리하고 모니터링하는 데 사용된다. 관리도는 샘플 공정의 평균 및 분산과 같은 샘플 통계를 플로팅(plottin)하여 품질 특성을 제공한다. 공정이 통제 불능 상태인 경우, 조기에 감지할 수 있으므로 제조업체는 공정 또는 기계 상태를 확인하여 시스템을 수정하기 위한 조치를 취할 수 있다. 이를 통해 최종 제품의 결함이 줄어들고 품질이 향상된다.

[0093] 본 발명에서는  $X$  및  $R$  관리도를 사용하여 공정을 모니터링한다. 두 차트 모두 SPC에서 일반적으로 사용되며, 생산 공정 품질을 측정하기 위해 별도로 사용된다. 공정 품질을 제어하기 위해서는 제어 상한(upper control limit, UCL), 중심선(center line, CL) 및 제어 하한(lower control limit, LCL)의 3 가지 직선을 계산해야 한다.  $X$  차트는 공정이 대상을 중심으로 하는 경우, 공정의 평균값 또는 범위 값을 나타낸다. 공정 산포가 올바른 범위 내에 있는 경우,  $R$  관리도는 생산 공정의 표준 편차를 나타낸다.

[0094]  $X$  차트는 다음 공식으로 계산된다.

$$UCL_X = \bar{\bar{x}} + A_2\bar{R}$$

$$CL_X = \bar{\bar{x}} = \frac{\sum_{j=1}^k \bar{x}_j}{k}, \text{ where } \bar{x}_j = 1/n_j \sum_{i=1}^{n_j} x_i$$

$$LCL_X = \bar{\bar{x}} - A_2\bar{R} \quad (7)$$

[0095] 여기서,  $x_i$ 는 실제 데이터,  $k$ 는 생산 공정의 그룹 크기,  $n_i$ 는  $i$  번째 그룹의 실제 데이터 수,  $\bar{R}$  는 모든 관측된 샘플 범위의 평균 값이며,  $\bar{x}_j$  는  $j$  번째 그룹의 모든 샘플의 평균이고,  $A_2$ 는 샘플 크기에 따라 그 값이 달라지는 계수이다.

[0097] 공정이 관리 한계를 벗어나면 비정상적인 공정이 발생했다는 통계적 확실성이 있으며 이는 제품 결함의 징후이다.  $R$  관리도는 샘플 생산 공정 범위를 사용하여 생산 공정의 스프레드 변화를 모니터링한다.  $i_1$ 이  $i$  번째 그룹에서 실제 데이터의 첫 번째 인덱스(index)인 경우, 그 값은 다음 공식에 의해 계산된다.

$$UCL_R = D_4\bar{R} \quad (8)$$

$$R_i = \text{Max}(x_{i_1}, \dots, x_{i_1+n_i}) - \text{Min}(x_{i_1}, \dots, x_{i_1+n_i}) \quad (9)$$

$$CL_R = \bar{R} = \frac{\sum_{i=1}^k R_i}{k} \quad (10)$$

$$LCL_R = D_3\bar{R} \quad (11)$$

[0102] 여기서,  $D_3$  및  $D_4$ 는 생산 공정 크기의 샘플에 따라 값이 결정되는 계수이며,  $\bar{R}$  는 모든 관측된 샘플 범위의 평균값이고,  $K$ 는 서브그룹의 수이다.  $R$  및  $X$  차트의 LCL 값이 0이면 표시되지 않고, 수평축과 수평을 이룬다.

- [0104] 이제 공정 능력 관리 분석(Process Capability Control Analysis, PCCA)에 대해 설명하면 다음과 같다.
- [0105] PCC(Process Capability Control)는 잘 알려져 있으며 제조 공정 성능을 측정하기 위한 선택 기준으로 널리 사용되며, 상당히 낮은 비용과 높은 생산성으로 허용 가능한 수준의 결과를 달성했다.
- [0106] PCCA(Process Capability Control Analysis)는 생산 공정 품질을 분석하는 데 사용되므로 제조업체는 공정 품질을 실시간으로 확인할 수 있다. 또한 공정 능력(Capability, Cp) 및 공정 능력(process capability, Cpk) 지수는 PCCA의 결과를 공식화하며, 제품 설계 사양 한계를 통해 생산 품질을 개선하기 위해 제조 산업에서 주로 사용된다.
- [0107] 제조업체는 생산 공정의 동작과 기계의 성능을 관찰하기 위해 사양 제한을 설정한다. 이러한 한계는 생산 공정이 제품을 생산하는 능력을 보여주며 공정이 사양 요구 사항 내에서 작동하는지 여부를 결정할 수 있다.
- [0108] 도 5는 공정 능력 관리를 분석한 그래프이다.
- [0109] 도 5는 규격 하한(Lower specification limit, LSL)과 규격 상한(Upper specification limit, USL)에 의해 결정된 PCCA의 결과를 보여준다. 사양 제한은 제조업체의 기능에 따라 각 기계마다 설계된다.
- [0110] 도 5의 그래프에서 평균  $\mu$ 는 공정 조건의 중앙값이고,  $\tau$ 는 하루의 목표 제품 수이다. Cp는 규격 한계 및 표준 편차를 사용하여 계산되는 반면, Cpk는 규격 한계, 표준 편차 및 공정 평균을 사용하여 계산된다. 이러한 것들은 다음 공식으로 계산된다.
- [0111] 
$$Cp = \frac{USL - LSL}{6\sigma} \quad (12)$$
- [0112] 식 12는 Cp를 계산하는 식을 보여준다. 여기서,  $\sigma$ 는 공정 조건 값의 표준 편차이다. 공정의 평균값이 규격 한계를 가진 경우, 계산된 결과를 통해 공정이 실현 가능한지 여부를 확인할 수 있습니다.
- [0113] 도 6은 공정 능력(Cp)의 기준을 보여주는 도표이다.
- [0114] 
$$Cpk = (1 - k) * cp \quad (13)$$
- [0115] 식 13은 Cpk를 계산하는 식이다. 여기서, k는 다음 식 14를 사용하여 계산되며, Cp와 Cpk 사이의 관계를 나타낸다. 그리고, u는 기계의 목표 실제 데이터의 평균이다.
- [0116] 
$$k = \frac{\frac{(USL + LSL)}{2} - u}{\frac{(USL - LSL)}{2}} \quad (14)$$
- [0117] Cpk의 계산 결과는 공정의 평균값이 규격 한계의 중앙값에서 벗어날 때 공정이 실현 가능한지 여부를 결정하는데 사용된다. Cpk는 공정 평균이 사양 한계 사이에 있지 않을 때를 고려한다.
- [0118] 도 7은 공정 능력 지수(Cpk)의 기준을 보여주는 도표이다.
- [0119] 도 6 및 도 7에서 볼 수 있는 바와 같이, Cp 및 Cpk의 계산 결과가 1보다 작으면, 공정을 실행할 수 없고 설계 사양을 충족하지 않는다. 반대로, Cp 및 Cpk의 계산 결과가 1보다 크거나 같으면, 공정이 실현 가능하고 설계 사양을 충족시킨다. 그리고, Cpk는 Cp보다 클 수 없으며, 실제 공정 평균이 설계 사양의 중간에 해당하는 경우 동일한 값을 가질 수 있다.
- [0121] 이제 본 발명에서 제조 공정 품질 예측에 대해 설명하고자 한다. 즉, 공정 품질의 예측 단계에 분석 데이터의 전처리 후 LSTM(long short term memory)을 적용하여 생산 공정 데이터에 대한 복합 통계적 분석 결과를 설명한다.
- [0122] 도 8은 본 발명의 일 실시예에 따른 공정 품질 성능을 결정하기 위해 LSTM을 적용하는 절차를 도시한 것이다.
- [0123] 도 8에서 보는 바와 같이, 전처리 후 4 가지 복합 통계적 분석 결과인 MCTA, PTAD, SPCA 및 PCCA가 기술된다. 또한, 전처리 결과는 딥스 레이어(dense layer)의 입력으로서 사용될 것이다. 딥스 레이어(dense layer)의 출력은 변형 레이어(reshape layer), LSTM 셀, 딥스 레이어(dense layer) 및 최종적으로 소프트맥스(Softmax) 레이어를 통과하여 공정 품질 성능을 결정한다.
- [0124] 도 2에 표시된 바와 같이, 본 발명에서는 PPD를 수집한다. 그리고, PPD 데이터에 복합 통계적 분석 방법을 적용

하여 네 가지 유형의 데이터로 변환하며, 이를 도표로 정리하면 다음과 같다.

- [0125] 도 9는 본 발명의 일 실시예에 따른 PPD 데이터에 복합 통계적 분석을 적용한 결과를 보여주는 도표이다.
- [0126] 도 9에 예시된 복합 통계적 분석 결과는 제조 공정 품질(MPQ)을 예측하기 위해, LSTM을 적용하기 전에 다음과 같이 전 처리된다.
- [0127] 범주형 데이터 인코딩(Encoding categorical data): 레이블 형 인코더와 하나의 핫 인코더가 범주형 데이터를 숫자 형 데이터로 변환하는 데 사용된다. 범주형 데이터는 숫자 값이 아닌 레이블 값을 포함하는 변수이다. 많은 머신 러닝 및 딥 러닝 알고리즘은 모든 입력 및 출력 변수가 숫자여야하므로 범주형 데이터는 숫자 값으로 변환되어야 한다.
- [0128] 데이터세트 분할(Splitting dataset): 데이터 세트는 트레이닝, 유효성 검사 및 테스트 세트의 세 가지 하위 집합으로 구분된다. 트레이닝 세트는 모델을 트레이닝하는 데 사용되는 실제 데이터이다. 유효성 검사 세트는 하이퍼 파라미터를 조정하고 트레이닝 데이터 세트에 주어진 모델을 평가하는 데 사용되는 데이터 세트이다. 테스트 세트는 모델이 완전히 트레이닝되어 최종 모델이 트레이닝 세트에 어떻게 부합되는지 평가하기 위해 사용되는 데이터 세트이다.
- [0129] 기능 스케일링(Feature scaling): 이 단계는 데이터 정규화라고도 하며 독립 변수 또는 데이터 기능의 범위를 정규화하는 데 사용된다. Z- 점수 표준화를 적용하여 값을 스케일링하는 데 기능 표준화가 사용된다. 이 방법은 머신 러닝 및 딥 러닝 모델의 정규화에 널리 사용되며, 각 기능에 대한 분포 평균 및 표준 편차를 계산한다. 그 다음 각 기능에서 평균을 뺀다. 그 다음 각 기능의 값을 표준 편차로 나눈다.
- [0130] LSTM은 시퀀스 데이터의 패턴을 인식하도록 설계된 강력한 유형의 인공 반복 신경망(RNN)이다.
- [0131] RNN은 입력 데이터를 기억하여 이 모델이 다음 순서를 예측할 수 있게 한다. 그러나 RNN은 단기 종속성만 처리할 수 있지만, LSTM은 장기 종속성을 학습할 수 있다.
- [0132] LSTM은 셀(cell), 입력 게이트(input gate), 출력 게이트(output gate) 및 망각 게이트(forget gate)로 구성된다.
- [0133] 셀은 임의의 시간 간격에 대한 값 기억을 처리한다. 입력 게이트는 새로운 입력을 허용할지 여부를 결정하고, 망각 게이트는 중요하지 않기 때문에 정보를 삭제할지 또는 출력 게이트의 현재 시간 단계에서 출력에 영향을 미치도록 할지 여부를 결정한다.
- [0134] LSTM 게이트는 아날로그이므로 시그모이드(sigmoid) 형태로 역 전파를 수행 할 수 있다. 이 아키텍처의 값 범위는 0에서 1까지이다. 이 아키텍처는 RNN의 한계를 극복하기 위해 장기간에 걸쳐 역사적 데이터를 기억하고 잊어버리는 능력을 향상시켜 설계되었다. LSTM은 시간과 계층을 통해 역프로포게이팅(backpropagated)될 수 있는 오류를 보존한다. LSTM은 더 지속적인 오류를 유지할 수 있기 때문에, RNN이 많은 시간 단계를 거쳐 계속 학습할 수 있도록 돕는다. 따라서 원인과 효과를 원격으로 연결하는 채널을 처리한다.
- [0135] LSTMs는 셀 게이트에 반복되는 네트워크의 정상적인 흐름 외부에 정보를 저장한다. 셀 게이트는 정보를 읽고 쓰고 지울 때 저장해야 할 정보와, 어떤 게이트가 열리고 닫히는지 결정한다. 셀 게이트는 0부터 1까지의 시그모이드 곱셈으로 구현된다. LSTM의 셀 게이트는 수신하는 신호에 대해 작동하며, 이는 신경망의 노드와 비슷하고, 정보를 차단하거나 전달할 수 있다.
- [0136] LSTM의 셀 게이트들은 필터에 사용할 수 있는 그들만의 가중치 세트를 가지고 있으며, 그들의 가중치는 반복적인 네트워크의 학습 과정에 의해 조정되는 변조 입력과 숨겨진 상태들이다. 경사 강하 및 역추진 오류의 조정 가중치에 의해 입력, 이탈 또는 삭제하기 위해 열린 정보가 있을 때 학습하는 것이 LSTM 셀이다. LSTM은 게이트라고 불리는 구조를 따라 정보를 제거하거나 셀 상태에 추가할 수 있다. 게이트는 시그모이드 함수와 점층적인 곱셈 연산으로 구성된다.
- [0137] 시그모이드 함수는 0에서 1까지의 결과를 출력한다.  $\sigma$ 는 시그모이드 활성화 함수이다. 게이트 활성화를 계산하기 위해 입력 및 가중치의 로지스틱 시그모이드 기능이 사용된다. 가중치는 각 시퀀스를 통해 LSTM 장치에 의해 분배된다. 셀 및 숨겨진 상태는 입력 및 출력 게이트의 제어를 받는다.
- [0138] 각 시퀀스의 블록 입력, 입력, 출력, 셀, 숨김 및 망각 상태는 각각  $z$ ,  $i$ ,  $o$ ,  $c$ ,  $h$  및  $f$ 로 표시된다. 그리고, 가중치 세트는  $\mathbb{W}$ 로 표시되며, LSTM 장치의 입력과 각 상태 사이를 연결한다. 출력 상태는 이전 시간 단계에서 전파된 정보를 관리한다. 망각 게이트를 포함하는 새로운 LSTM 아키텍처가 소개되었는데, 망각 게이트는 LSTM

메모리 셀이 하나의 시퀀스 학습을 마치고 새 시퀀스를 시작하기 전에 재시작 할 수 있다. LSTM 단위 방정식은 다음과 같이 나타낼 수 있다.

[0139] 
$$z_t = \tanh(W^z x_t + R^z h_{t-1} + b^z) \quad (15)$$

[0140] 
$$i_t = \sigma(W^i x_t + R^i h_{t-1} + b^i) \quad (16)$$

[0141] 
$$f_t = \sigma(W^f x_t + R^f h_{t-1} + b^f) \quad (17)$$

[0142] 
$$o_t = \sigma(W^o x_t + R^o h_{t-1} + b^o) \quad (18)$$

[0143] 
$$c_t = z_t \odot i_t + c_{t-1} \odot f_t \quad (19)$$

[0144] 
$$h_t = \tanh(c_t) \odot o_t \quad (20)$$

[0145] 여기서  $x_t$ 는 현재 입력 시간 단계이고,  $s_t$ 는 시그모이드 함수이고,  $\odot$ 는 포인트 단위 연산을 나타내고,  $W^*$ 는 입력 가중치이고,  $R^*$ 는 반복 가중치이고,  $b$ 는 바이어스이다.

[0146] 생산 공정 품질의 다음 순서를 예측하기 위하여, 통계적 공정 분석 결과가 가장 중요한 특징이며, 이러한 특징은 LSTM 모델에서 향후 공정 품질을 트레이닝하고 예측하는 데 사용된다.

[0147] 본 발명에서 예측 과정은 다음과 같이 두 단계로 구분된다.

[0148] LSTM 모델을 설립하기 전에 범주형 데이터 인코딩, 데이터 세트 분할 및 기능 스케일링과 같은 데이터 전 처리가 구현된다. 이 단계에서는 범주형 데이터 인코딩 또는 원핫(one-hot) 인코딩을 사용하여 데이터 세트의 범주형 변수를 예측 모델에서 더 나은 작업을 수행하기 위해 제공하는 형식으로 변환한다. 그 다음, 데이터 세트를 트레이닝, 유효성 검사 및 테스트 세트들로 구분한다. 그 다음, 모든 기능 값을 표준화 된 값으로 변환한다. 이 값의 평균은 0이고 표준 편차는 1이다. 그리고, 다음과 같이 다양한 레이어로 구성된 신경망이 생성된다.

[0149] 입력 텐스 레이어(Input dense layer)는 신경망의 첫 번째 계층이다. 트레이닝 과정 동안, 데이터 세트는 신경망의 입력으로 전달된다. 입력 데이터 세트는 2 차원 배열이다. 이 레이어는 전과 트레이닝 절차를 사용하여 학습되고, 이 절차는 시그모이드(sigmoidal) 또는 소프트맥스(Softmax) 기능과 같은 활성화 기능에 의해 계산되는 뉴런 출력으로 구성된다. 텐스 레이어의 각 뉴런은 많은 입력과 하나의 출력을 가지고 있고, 뉴런의 수는 출력 수와 같다.

[0150] 형태 변경 레이어(Reshape layer)는 1 차원 또는 2 차원 배열일 수 있는 시퀀스 데이터를 LSTM 입력 계층의 필요한 3 차원 배열로 변환하는데 사용된다. 이는 예측 과정에서 모델을 맞추기 위하여, LSTM의 입력 레이어가 샘플, 시간 단계 및 기능과 같은 단일 값을 포함하는 3 차원 배열을 예상한다는 것을 의미한다.

[0151] LSTM 셀 레이어(LSTM cell layer)는 가장 많은 계산이 수행되는 레이어로, 전체 신경망에서 수행되며 RNN 레이어에 유지된다. RNN 계층은 LSTM 셀로 구성되고, 출력 계산에는 여러 가지 방법이 있다. 각 신경 레이어의 각 뉴런의 출력은 입력으로 전달된다. 이를 통해 신경 레이어 수를 줄이고, 네트워크가 트레이닝되는 동안 에포크(epoch) 수를 제한하여 신경 네트워크의 트레이닝 공정 속도를 높이고, 의미있는 예측을 제공할 수 있다.

[0152] 출력 텐스 레이어(Output dense layer)는 RNN 계층에서 유지되는 전체 신경망에 의해 수행되는 계층이다. RNN 계층은 LSTM 셀로 구성되고, 출력 계산에는 여러 가지 방법이 있습니다. 각 신경 레이어의 각 뉴런의 출력은 다음 레이어의 입력으로 전달됩니다. 이를 통해 신경 레이어 수를 줄이고, 네트워크가 트레이닝되는 동안 에포크 수를 제한하여 신경 네트워크의 트레이닝 공정 속도를 높이고, 의미있는 예측을 제공할 수 있다.

[0154] 소프트맥스(Softmax) 함수 알고리즘은 LSTM 예측 모델의 마지막 계층에서 사용된다. 이 알고리즘을 사용하여 LSTM 출력 레이어의 예측 결과를 분류한다. 소프트맥스(Softmax) 함수는 K 실수 벡터를 입력으로 받아서 K 확률로 구성된 확률 분포로 정규화하는 활성화 알고리즘이다. 즉, 소프트맥스(Softmax) 함수를 적용하기 전에 일부 벡터 구성 요소는 음수이거나 1보다 클 수 있으며, 1의 합이 아닐 수 있다. 그러나 소프트맥스(Softmax) 기능을 적용한 후 각 구성 요소는 간격 (0, 1)이 되고 구성 요소는 최대 1이 되어 확률로 해석 할 수 있다. 또한 입력 요소가 클수록 확률이 커진다. 소프트맥스(Softmax) 함수는 신경망에서 종종 정규화되지 않은 네트워크 출력을 예측된 출력 클래스에 대한 확률 분포에 매핑(mapping)하는 데 사용된다. 소프트맥스(Softmax) 분류는 이진 형

태의 로지스틱 회귀(logistic regression)의 일반화이다. 회귀 출력이 정수 또는 부동값과 같은 숫자에도 불구하고, 분류는 주어진 입력에 대한 다양한 클래스를 예측하는 감독된 기계학습의 문제다. 이러한 수량은 양과 크기와 같은 경우가 많다. 본 발명에서 다음과 같은 공식으로 소프트맥스(Softmax) 함수를 사용하여 MPQ 결과를 출력한다.

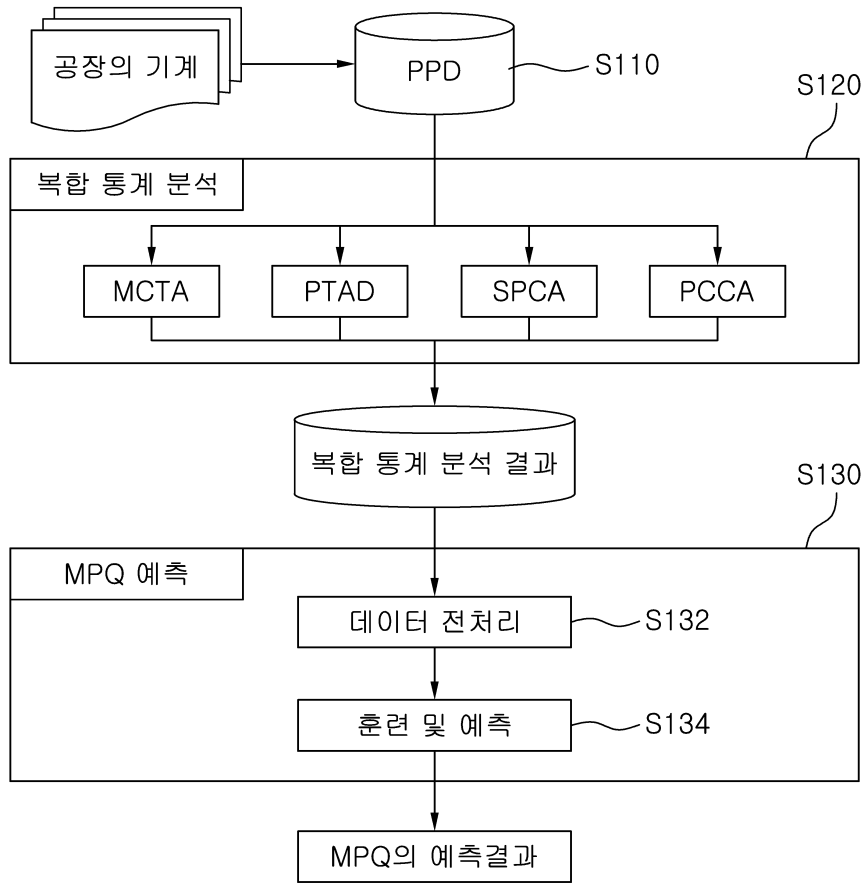
$$MPQ_i = \frac{e^{a_i}}{\sum_{j=1}^N e^{a_j}} \quad (21)$$

- [0155]
- [0156] 이 알고리즘을 사용하여 LSTM 출력 레이어의 예측 결과를 분류한다. 여기서 MPQ<sub>i</sub>는 크기 N의 입력 벡터에서 i 번째 값에 대한 소프트맥스(Softmax) 출력이다. 입력 벡터는 [h<sub>1</sub>, h<sub>2</sub>, h<sub>3</sub>, ..., h<sub>n</sub>]이고, MPQ<sub>i</sub>는 항상 양수이다. 표준 지수 함수는 각 요소 MPQ<sub>i</sub>에 적용되며, 그 값은 모든 지수의 합으로 나누어 정규화된다.
- [0157] 이 정규화는 출력 벡터 MPQ의 구성 요소의 합이 1이 되도록 한다. 지수때문에 MPQ<sub>i</sub>가 0보다 크다. 분자가 분모에 나타나고 다른 양수와 합산하기 때문에 MPQ<sub>i</sub>는 1보다 작다.
- [0158] MPQ 상태에 대한 결정 기준은 MPQ 값에 따라 기계 실행 가능성 및 작동 상태 조건을 확인하여 많은 실험을 통해 결정되었다. 도 10은 생산 공정 안정성 기준을 보여준다.
- [0159] 도 10은 MPQ의 기준을 예시한 도표이다.
- [0160] 도 10을 참조하면, 소프트맥스(Softmax) 분류기 결과가 0.50에서 1.00 사이인 경우가 최상의 결과이며, 이는 공정이 결함없이 제품을 생산할 수 있음을 의미한다.
- [0161] 반면, 분류기 결과가 0.00에서 0.50 사이인 경우, 도 10의 기준 표에 의하면 품질이 좋지 않거나 최악이다. 이는 공정이 너무 나빠서 결함있는 제품의 생산을 방지하기 위해 가능한 한 빨리 개선해야 한다는 것을 나타낸다.
- [0162] 평가 매트릭(evaluation metric)은 학습된 모델의 적용 가능성을 측정하고, 모델의 성능을 설명한다. 이를 통해 학습한 모델이 원하는 정확도를 달성할 때까지 개선하고, 계속할 수 있다.
- [0163] 모델의 예측 정확도에 대한 다양한 평가 매트릭이 있다. 평가 매트릭 선택은 모델 유형과 구현 계획에 따라 달라진다. 여기서는 혼동 매트릭스(confusion matrix)를 사용하여 예측 모델을 평가한다. 혼동 행렬은 분류 모델의 성능을 측정하는 데 사용되는 고도 매트릭이다.
- [0164] 도 11은 혼동 매트릭스의 네가지 주요 요소를 보여주는 도표이다.
- [0165] 도 11의 혼동 매트릭스 표를 참조하면, 예측 분석에서 가 양성(false positive), 가 음성(false negative), 진 양성(true positive) 및 진 음성(true negative)의 수를 표시하는 2 개의 행과 2 개의 열이 있다. 혼동 매트릭스 표에서 행렬의 각 행은 예측된 클래스를 나타내고 각 열은 실제 클래스를 나타낸다.
- [0166] 도 12는 혼동 매트릭스의 오류 매트릭을 나타낸 도표이다.
- [0167] 도 12를 참조하면, 혼동 매트릭스는 다양한 접근 방식에서 분류 모델의 성능을 측정하는 데 사용할 수 있다. 도 12의 표는 정확도(Accuracy), 민감도(Sensitivity), 특이성(Specificity), 정밀도(Precision) 및 리콜(Recall)과 같은 혼동 매트릭스의 특정 오류 매트릭 계산을 보여준다. 민감도는 올바르게 예측된 양성 클래스의 비율에 대응되는 것인 반면, 특이성은 실수로 양성으로 간주되는 음성 클래스의 비율에 대응되며, 둘 다 0에서 1 사이의 값을 갖는다.
- [0168] 한편, 본 발명의 실시예에 따른 제조 공정 품질 분석 및 예측 방법은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현되는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다.
- [0169] 예컨대, 컴퓨터가 읽을 수 있는 기록매체로는 롬(ROM), 램(RAM), 시디-롬(CD-ROM), 자기 테이프, 하드디스크, 플로피디스크, 이동식 저장장치, 비휘발성 메모리(Flash Memory), 광 데이터 저장장치 등이 포함된다.
- [0170] 또한, 컴퓨터로 읽을 수 있는 기록매체는 컴퓨터 통신망으로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 읽을 수 있는 코드로서 저장되고 실행될 수 있다.
- [0171] 이상 본 발명을 몇 가지 바람직한 실시예를 사용하여 설명하였으나, 이들 실시예는 예시적인 것이며 한정적인 것이 아니다. 본 발명이 속하는 기술분야에서 통상의 지식을 지닌 자라면 본 발명의 사상과 첨부된 특허청구범

위에 제시된 권리범위에서 벗어나지 않으면서 다양한 변화와 수정을 가할 수 있음을 이해할 것이다.

도면

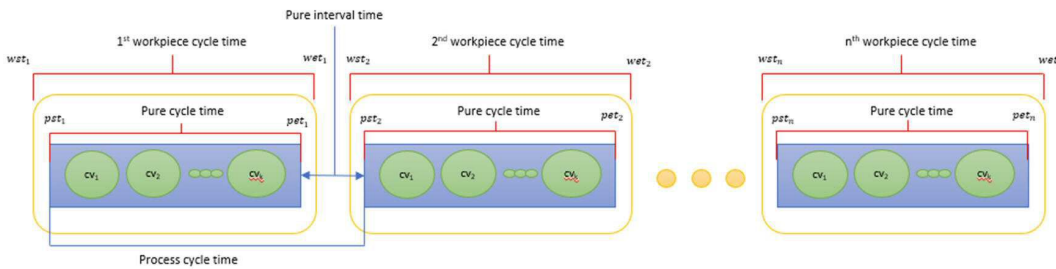
도면1



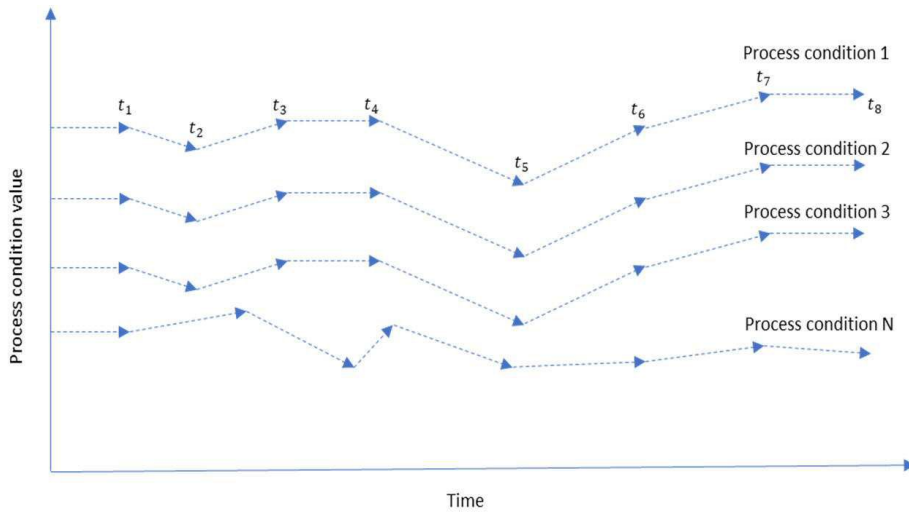
도면2

| Data description           | Data name                           | Description  |
|----------------------------|-------------------------------------|--|
| Machine                    | Machine_name                        | A machine to produce products                                      |
| Product                    | Product_name                        | Product name produced by a machine                                 |
| Process condition variable | PCV_name                            | Name of a condition variable related to a product                  |
| Workpiece start time       | Workpiece_start_time                | The time when a machine starts a single product                    |
| Workpiece end time         | Workpiece_end_time                  | The time when a machine finishes the product                       |
| Process start time         | Process_start_time_name             | Start time of the condition variable process                       |
| Process end time           | Process_end_time_name               | End time of the condition variable process                         |
| Real data                  | Real_data                           | Gathering real data during the condition variable processing       |
| Upper specification limit  | Upper specification limit(USL)_name | Upper specification limit for the condition variable processing    |
| Lower specification limit  | Lower specification limit(LSL)_name | Lower specification limit for the condition variable processing    |
| Result of product          | Result                              | Product result (OK or NG) with real data of the condition variable |

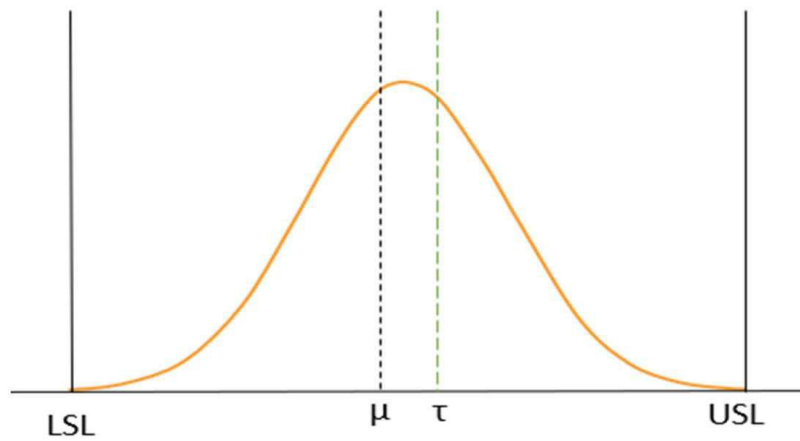
도면3



도면4



도면5



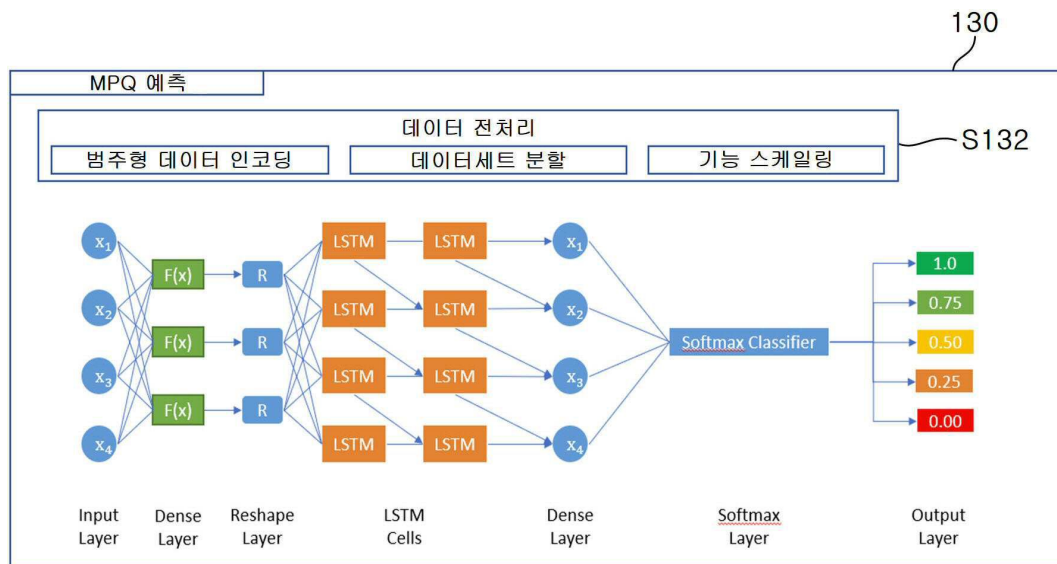
도면6

| Range                     | Quality   | Normal production rate | Failure production rate |
|---------------------------|-----------|------------------------|-------------------------|
| $C_p \geq 1.68$           | excellent | 99.999991%             | 0.000009%               |
| $1.34 \leq C_p \leq 1.67$ | very good | 99.999943%             | 0.000057%               |
| $1.01 \leq C_p \leq 1.33$ | good      | 99.9937%               | 0.0063%                 |
| $0.68 \leq C_p \leq 1.00$ | fair      | 99.73%                 | 0.27%                   |

도면7

| Range       | Quality   | Normal production rate | Failure production rate |
|-------------|-----------|------------------------|-------------------------|
| Cpk < 0.33  | terrible  | 99.85%                 | 1.15%                   |
| 0.33 ~ 0.67 | very poor | 95.45%                 | 4.65%                   |
| 0.67 ~ 1.00 | poor      | 99.73%                 | 0.27%                   |
| 1.00 ~ 1.33 | average   | 99.9937%               | 0.0063%                 |
| 1.33 ~ 1.67 | good      | 99.999943%             | 0.000057%               |
| 1.67 ~ 2.00 | very good | 99.999991%             | 0.000009%               |
| Cpk > 2.00  | excellent | 99.999998%             | 0.000002%               |

도면8



도면9

| Classification                        | Outcomes                                | Data type |
|---------------------------------------|---|-----------|
| Manufacturing cycle time              | Stability of workpiece cycle time(sWCt) | long      |
|                                       | Stability of pure cycle time(sPrCt)     | long      |
|                                       | Stability of pure interval time (sPrIt) | Long      |
|                                       | Stability of process cycle time (sPCt)  | Long      |
| Process trajectory abnormal detection | Abnormal process status value (APSV)    | Float     |
| Statistical process control           | Average of a X-Chart(X-AVG)             | Float     |
|                                       | Upper control limit of X-Chart (X-UCL)  | Float     |
|                                       | Lower control limit of X-Chart(X-LCL)   | Float     |
|                                       | centerlineof X-Chart(X-CL)              | Float     |
|                                       | Average of R-Chart(X-AVG)               | Float     |
|                                       | Upper control limit of R-Chart(R-UCL)   | Float     |
|                                       | Lower control limit of R-Chart(R-LCL)   | Float     |
| Process Capability                    | Process capability (Cp)                 | Float     |
|                                       | Process capability index (Cpk)          | Float     |

도면10

| MPQ value   | Quality   |
|-------------|-----------|
| 1.00        | Excellent |
| 0.75 - 0.99 | Good      |
| 0.50 - 0.75 | Fair      |
| 0.25 - 0.50 | Poor      |
| 0.00 - 0.25 | Terrible  |

도면11

|              |          | Predicted classes   |                     |
|--------------|----------|---------------------|---------------------|
|              |          | Positive            | Negative            |
| Actual Class | Positive | True positive(TP)   | False negative (FN) |
|              | Negative | False positive (FP) | True negative (TN)  |

도면12

| Error Metric | Definition                      |
|--------------|---------------------------------|
| Accuracy     | $TP + TN / (TP + TN + FP + FN)$ |
| Sensitivity  | $TP / (TP + FN)$                |
| Specificity  | $TN / (TN + FP)$                |
| Precision    | $TP / (TP + FP)$                |
| Recall       | $TP / (TP + FN)$                |