



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년06월27일
 (11) 등록번호 10-1871940
 (24) 등록일자 2018년06월21일

(51) 국제특허분류(Int. Cl.)
G06F 19/00 (2018.01)

(21) 출원번호 10-2014-0056574

(22) 출원일자 2014년05월12일

심사청구일자 2017년02월07일

(65) 공개번호 10-2015-0129507

(43) 공개일자 2015년11월20일

(56) 선행기술조사문헌

KR1020140056952 A*

US20080215513 A1*

Fussy-Genetic Algorithm 기반의 자가적응형 돌발
 상황 검지모형 개발 연구”, 이시복 외 1인, 대
 한교통학회지 제22권 제4호, (2004.08.31.)*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

한화에어로스페이스 주식회사

경상남도 창원시 성산구 창원대로 1204 (성주동)

한양대학교 산학협력단

서울특별시 성동구 왕십리로 222(행당동, 한양대
 학교내)

(72) 발명자

조철형

경상남도 창원시 성산구 창원대로 1204 (성주동)

방현진

경상남도 창원시 성산구 창원대로 1204 (성주동)

(뒷면에 계속)

(74) 대리인

리엔특허법인

전체 청구항 수 : 총 6 항

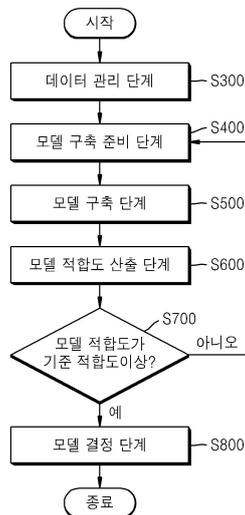
심사관 : 김병균

(54) 발명의 명칭 **설비 이상 예측 모델의 구축방법 및 시스템**

(57) 요약

본 발명은 설비 데이터와 비선형관계를 가지는 설비 이상을 실시간으로 진단하기 위한 방법을 개시한다. 이를 위해 과거의 설비 이상을 포함하는 설비 데이터를 입력받고, 유전적 알고리즘으로 설비 이상 예측 모델을 구축하기 위한 변수를 생성한 뒤 설비 이상 예측 모델을 구축 및 검증한다. 미리 설정된 기준 적합도를 만족할 때까지 설비 이상 예측 모델을 구축하기 위한 변수 생성과 모델의 구축, 검증을 반복하며 최종 결정된 설비 이상 예측 모델로 설비 데이터로부터 설비 이상을 실시간으로 진단할 수 있다.

대표도 - 도3



(72) 발명자

김경택

경상남도 창원시 성산구 창원대로 1204 (성주동)

김덕수

서울특별시 성동구 왕십리로 222 (사근동)

이기찬

서울특별시 성동구 왕십리로 222 (사근동)

조영송

서울특별시 성동구 왕십리로 222 (사근동)

명세서

청구범위

청구항 1

공정마다 생산물을 생산하는 설비의 설비 데이터를 입력받고 동기화하는 데이터 관리 단계;
 최초 크로모솜을 정의하는 모델 구축 준비 단계;
 상기 동기화된 설비 데이터를 학습 데이터와 시험 데이터로 나누는 데이터 분할 단계;
 상기 최초 크로모솜으로부터 상기 학습 데이터를 모두 반영하는 예측 모델을 구축하는 모델 구축 단계;
 상기 예측 모델에 상기 시험 데이터를 입력하여 모델 적합도를 산출하는 모델 적합도 산출 단계;
 미리 설정된 기준 적합도보다 상기 모델 적합도가 높거나 같은 경우 상기 예측 모델을 설비 이상 예측 모델로 결정하는 모델 결정 단계; 및
 미리 설정된 기준 적합도보다 상기 모델 적합도가 낮은 경우 현재의 크로모솜을 전세대 크로모솜으로 새로 정의하고, 상기 전세대 크로모솜을 교차 및 돌연변이 시켜 만든 후세대 크로모솜으로 새로운 예측 모델을 구축하고, 상기 모델 적합도 산출단계부터 반복 진행하는 모델 구축 반복 단계;를 포함하고,
 상기 데이터 관리 단계는,
 상기 설비가 설비이상 발생 공정을 통해 생산된 생산물을 폐기하면, 상기 폐기된 생산물의 양을 기초로 산출된 누적버림률(accumulated loss rate)를 상기 동기화된 설비 데이터에 포함시키고,
 상기 설비 이상 예측 모델은 입력된 새로운 설비 데이터에 포함된 누적버림률을 기초로 상기 설비의 이상동작시점을 예측하고,
 상기 기준 적합도는,
 허위양성(false positive) 및 커버리지(coverage)로 구성된 것을 특징으로 하는 설비 이상 예측모델 구축방법.

청구항 2

제1항에 있어서, 상기 예측 모델은
 서포트 벡터 머신(SVM : Support Vector Machine) 모델을 이용하는 것을 특징으로 하는 설비 이상 예측 모델 구축 방법.

청구항 3

제2항에 있어서,
 상기 최초 크로모솜, 전세대 크로모솜 및 후세대 크로모솜은,
 이진수(binary number)로 표현된 입력변수의 조합과 커널함수 파라미터를 포함하는 것을 특징으로 하는 설비 이상 예측 모델 구축 방법.

청구항 4

제1항 내지 제3항 중 어느 하나에 있어서,
 상기 설비 이상과 상기 설비 데이터는 상호 비선형적 관계를 가지는 것을 특징으로 하는 설비 이상 예측 모델 구축 방법.

청구항 5

제1항에 있어서,
 상기 설비 이상 예측 모델로 실시간으로 유입되는 설비 데이터로부터 설비 이상을 진단하는 설비 이상 예측 모

텔 구축 방법.

청구항 6

공정마다 생산물을 생산하는 설비의 설비 데이터를 입력받고 동기화하는 데이터 관리부;

최초 크로모즘을 정의하는 최초 크로모즘 정의부;

상기 동기화된 설비 데이터를 학습 데이터와 시험 데이터로 나누는 데이터 분할부;

상기 최초 크로모즘으로부터 상기 학습 데이터를 모두 반영하는 예측 모델을 구축하는 예측 모델 구축부;

상기 예측 모델에 상기 시험 데이터를 입력하여 모델 적합도를 산출하는 모델 적합도 생성부;

미리 설정된 기준 적합도와 상기 모델 적합도를 비교하는 적합도 비교부; 및

상기 기준 적합도보다 상기 모델 적합도가 높거나 같은 경우 상기 예측 모델을 설비 이상 예측 모델로 결정하고, 그 외에 경우는 현재의 크로모즘을 전세대 크로모즘으로 새로 정의하고, 상기 전세대 크로모즘을 교차 및 돌연변이 시켜 만든 후세대 크로모즘으로 새로운 예측 모델을 구축하고, 상기 모델 적합도 생성부의 동작부터 설비 이상 예측 모델이 결정될 때까지 반복 동작하는 후세대 크로모즘 정의부;를 포함하고,

상기 데이터 관리부는,

상기 설비가 설비이상 발생 시 발생한 공정을 통해 생산된 생산물을 폐기하면, 상기 폐기된 생산물의 양을 기초로 산출된 누적버림률(accumulated loss rate)을 상기 동기화된 설비 데이터에 포함시키고,

상기 설비 이상 예측 모델은 입력된 새로운 설비 데이터에 포함된 누적버림률을 기초로 상기 설비의 이상동작시점을 예측하고,

상기 기준 적합도는,

허위양성(false positive) 및 커버리지(coverage)로 구성된 것을 특징으로 하는 설비 이상 예측 모델 구축 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 설비 이상 예측 모델의 구축방법 및 시스템에 관한 것으로, 구체적으로는 설비데이터와 설비 이상간에 관계성 모델을 도출하고 이를 적용하여 설비데이터로부터 설비 이상을 실시간으로 진단하는 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 설비 데이터란 설비를 동작 및 관리하면서 산출되는 모든 데이터를 의미하며, 설비 이상은 상기 설비를 동작시키는 데 있어서 발생하는 외부 요인을 제외하고 설비 자체를 원인으로 하는 결함, 오동작을 의미한다.

[0003] 종래에는 선형성이 명확한 설비 데이터와 설비 이상을 정의하여 각종 통계적 방법을 사용하여 그 관계성을 찾아 관리해 왔다. 그러나, 설비 데이터와 설비 이상간의 관계가 반드시 선형성을 가지는 것은 아니고, 설비 데이터와 설비 이상간에 비선형성을 가지는 경우에 상기 통계적 방법(다변량, SPC, PCA)으로는 관계성을 찾기 어려워 설비 이상을 예측하기 곤란하고, 변경점이 발생했을 시에 대처가 어렵다는 점이 있었으며, 일반적인 설비데이터는 정규성과 등분산성을 보이지 않아 다양한 방식의 비모수 방법론들을 찾아야 하지만 시간이 많이 걸리고, 신뢰성을 높이기 어렵다는 한계가 있었다.

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 설비 이상과 선형성을 가지는 설비 데이터뿐만 아니라 비선형성을 가지는 설비 데이터로부터 설비 이상을 예측할 수 있는 관계성 모델을 제공하고 이를 활용하여 설비 이상을 실시간으로 예측하는 데에 있다.

과제의 해결 수단

[0005] 상기의 기술적 과제를 해결하기 위한 본 발명의 제 1실시예는, 설비 데이터를 입력받고 동기화하는 데이터 관리 단계; 최초 크로모즘을 정의하는 모델 구축 준비 단계; 상기 설비 데이터를 학습 데이터와 시험 데이터로 나누는 데이터 분할 단계; 상기 최초 크로모즘으로부터 상기 학습 데이터를 모두 반영하는 예측 모델을 구축하는 모델 구축 단계; 상기 예측 모델에 상기 시험 데이터를 입력하여 모델 적합도를 산출하는 모델 적합도 산출 단계; 미리 설정된 기준 적합도보다 상기 모델 적합도가 높거나 같은 경우 상기 예측 모델을 설비 이상 예측 모델로 결정하는 모델 결정 단계; 및 미리 설정된 기준 적합도보다 상기 모델 적합도가 낮은 경우 현재의 크로모즘을 전세대 크로모즘으로 새로 정의하고, 상기 전세대 크로모즘을 교차 및 돌연변이 시켜 만든 후세대 크로모즘으로 새로운 예측 모델을 구축하고, 상기 모델 적합도 산출단계부터 반복 진행하는 모델 구축 반복 단계; 를 포함한다.

[0006] 상기의 기술적 과제를 해결하기 위한 본 발명의 제 2실시예는, 입력받은 설비 데이터를 동기화하는 데이터 관리부; 최초 크로모즘을 정의하는 최초 크로모즘 정의부; 상기 설비 데이터를 학습 데이터와 시험 데이터로 나누는 데이터 분할부; 상기 최초 크로모즘으로부터 상기 학습 데이터를 모두 반영하는 예측 모델을 구축하는 예측 모델 구축부; 상기 예측 모델에 상기 시험 데이터를 입력하여 모델 적합도를 산출하는 모델 적합도 생성부; 미리 설정된 기준 적합도와 상기 모델 적합도를 비교하는 적합도 비교부; 및 상기 기준 적합도보다 상기 모델 적합도가 높거나 같은 경우 상기 예측 모델을 설비 이상 예측 모델로 결정하고, 그 외에 경우는 현재의 크로모즘을 전세대 크로모즘으로 새로 정의하고, 상기 전세대 크로모즘을 교차 및 돌연변이 시켜 만든 후세대 크로모즘으로 새로운 예측 모델을 구축하고, 상기 모델 적합도 생성부의 동작부터 설비 이상 예측 모델이 결정될 때까지 반복 동작하는 후세대 크로모즘 정의부;를 포함한다.

발명의 효과

[0007] 본 발명에 의해 도출된 모델을 이용하면 설비 데이터와 설비 이상간에 선형관계를 가질 때는 물론, 비선형관계를 가지더라도 설비 데이터로부터 설비 이상을 높은 확률로 예측할 수 있어서, 해당 설비를 관리하는 데에 도움을 줄 수 있다.

[0008] 구체적으로는, 설비 데이터의 경향성을 읽어내어 설비 이상을 조속히 판단함으로써, 더 큰 설비의 결함이나 오동작을 막고 사고를 미연에 방지할 수 있게 된다.

도면의 간단한 설명

[0009] 도 1은 종래의 선형성이 명확한 설비 데이터와 설비 이상을 통계적 방법을 사용하여 그 관계성을 찾는 방법을 나타낸다.

도 2는 본 발명에 따른 설비 이상 예측 모델 구축 시스템을 나타낸 블록도이다.

도 3은 본 발명에 따른 설비 이상 예측 모델의 구축 방법을 나타낸 순서도이다.

도 4은 데이터 관리 단계에서 일어나는 과정을 상세하게 나타낸 흐름도이다.

도 5는 본 발명의 실시예 중 하나인, 서포트 벡터 머신 모델을 설비 이상 예측 모델로 구축하는 과정을 나타내는 것으로, 특히 데이터 관리부를 제외한 모델 구축 준비부와 모델 구축부에서의 동작을 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0010] 이하 본 발명의 바람직한 실시예가 첨부된 도면들을 참조하여 설명될 것이다. 도면들 중 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 참조번호들 및 부호들로 나타내고 있음에 유의해야 한다.

[0011] 본 발명을 설명함에 있어서 관련된 공지 기능 혹은 구성에 대해 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그에 대한 상세한 설명은 생략하거나 간략하게 설명하는 것으로 한다. 한편, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0012] 도 1은 종래의 선형성이 명확한 설비 데이터와 설비 이상을 통계적 방법을 사용하여 그 관계성을 찾는 방법을 나타낸다.

- [0013] 정상관리선으로부터 벗어난 설비데이터 FAULT I내지 IV에 의해 설비 이상이 발생한 것을 개략적으로 나타내고 있다.
- [0014] 통계적으로 두 데이터의 상관성을 찾아내기 위해 분류 모델(Classification Model)을 사용한다. 분류 모델은 기계학습알고리즘에 의해 구축될 수 있는데, 기계학습 알고리즘이란 주어진 데이터를 컴퓨터에 입력하고 어떠한 특정 알고리즘을 기반으로 학습을 행하여 판별기준을 구축함으로써 새로운 데이터가 주어졌을 때 그 데이터가 어떠한 종류로 판별되는지를 예측하게 되는 과정을 가리킨다. 기계학습알고리즘에는 케이-근접이웃(KNN : K-Nearest Neighbors algorithm), 퍼셉트론(perceptron), 방사 기저 함수 네트워크(Radial Basis Function Network), 유전적 알고리즘(GA : Genetic Algorithm), 서포트 벡터 머신(SVM : Support Vector Machine) 등이 있으며 본 발명에서는 이 중 적어도 하나 이상의 알고리즘을 통해 설비 이상 예측 모델을 구축하는 것을 제안한다. 본 발명의 설비 이상 예측 모델이란, 설비 데이터를 입력받아서 설비 이상을 진단할 수 있는 모델을 의미하며 특히 본 발명의 실시예 중 하나에서는 유전적 알고리즘과 서포트 벡터 머신을 동시에 이용하고 있다.
- [0015] 유전적 알고리즘이란, 기계학습 알고리즘의 일종으로서 다윈의 진화론을 모티브로 하여 주위 환경에 따라 스스로 진화해나가는 방식으로 수행되는 알고리즘으로 기본적으로 1회 이상의 반복과정을 전제로 한다는 특징이 있다.
- [0016] 유전적 알고리즘의 진행은 여러 가지 선택법 중 하나의 선택법에 따라 복수 개의 크로모솜을 만들고, 각각의 크로모솜이 얼마나 우수한지 판별한 후, 우수한 크로모솜만 남기고 나머지 크로모솜을 제거하여 일정 기준을 넘는 크로모솜이 나올 때까지 진화를 거듭시켜 나가는 것이다. 이에 대한 자세한 설명은 도 5를 참고하여 후술한다.
- [0017] 서포트 벡터 머신은 두 가지 종류의 데이터를 적절하게 나누는 판별방식을 컴퓨터로 학습하여 새로운 데이터에 대한 예측을 수행하는 방식이다. 서포트 벡터 머신 모델을 이용하면 비선형분류라는 현실적인 문제들에서는 만족스런 성능을 낼 수 없다는 한계가 존재했으나 커널(kernel)이라는 함수를 사용한 맵핑(mapping) 방식을 적용하면 효과적으로 예측이 가능해지며, 커널을 통한 맵핑이라는 것은 우리가 실제로 데이터를 배치하는 입력공간에서는 잘 나누어지기 힘든 비선형문제를 특징 공간(feature space)라는 고차원의 공간으로 이동시켜서 이 새로운 공간에서 서포트 벡터 머신의 선형판별을 수행함으로써, 마치 처음의 입력 공간(input space)에서 복잡한 비선형 판별 문제를 해결한 것과 같은 효과를 얻는 것을 가리킨다.
- [0018] 서포트 벡터 머신 모델을 이용하기 위해서는 입력 벡터(input vector)를 결정하는 데 이용되는 입력 변수(input parameter)와 커널함수를 이용하기 위한 커널함수 파라미터(kernel function parameter), 이상 두 가지가 필요하다. 이에 대한 자세한 설명은 도 5를 참고하여 후술하며 상술한 서포트 벡터 머신은 공지 기술이므로 본 발명과 직접적으로 연관이 없는 일반적인 설명은 발명의 간명화를 위해 생략한다.
- [0019] 도 2는 본 발명에 따른 설비 이상 예측 모델 구축 시스템을 나타낸 블록도이다.
- [0020] 도 2를 참조하면, 본 발명에 따른 설비 이상 예측 모델 구축 시스템은 데이터 관리부(200), 모델 구축 준비부(210), 및 모델 구축부(220)를 포함한다.
- [0021] 데이터 관리부(200)는 설비 데이터를 관리하며 데이터 입력부(201)와 데이터 완성부(202)를 포함한다. 데이터 입력부(201)는 텍스트 파일형태의 생산 데이터 파일과 트레이스 데이터 파일을 읽고 저장한 뒤 이를 동기화한다. 데이터 완성부(202)는 누적버림률(accumulated loss rate)로 설비 이상을 나타내는 반응변수를 설정하고, 반응변수 자체를 생산 데이터에 포함시켜서 설비데이터를 완성시킨다. 즉, 완성된 설비 데이터는 과거에 설비가 정상적으로 동작할 때의 데이터 뿐만 아니라, 설비 이상이 발생했을 때의 데이터도 포함하게 된다. 완성된 동기화된 설비 데이터는 모델 구축 준비부(210)로 전달된다.
- [0022] 모델 구축 준비부(210)는 최적의 설비 이상 예측 모델을 구축하기 위한 변수를 유전적 알고리즘을 통해 찾으며 최초 크로모솜 정의부(211), 적합도 비교부(212), 및 후세대 크로모솜 정의부(213)을 포함한다.
- [0023] 최초 크로모솜 정의부(211)은 '설비 이상 예측 모델'을 결정하기 위해 실험적으로 만들어지는 '예측 모델'을 구축하는 변수인 최초 크로모솜을 정의한다. 최초 크로모솜이 포함하는 예측 모델의 구축 변수의 개수는 단수뿐만 아니라 복수도 될 수 있으며 그 예로서, 예측 모델을 서포트 벡터 머신 모델로 하는 본 발명의 실시예에 의한 경우 최초 크로모솜은 입력 변수의 조합과 커널함수 파라미터를 포함한다. 적합도 비교부(212)와 후세대 크로모솜 정의부(213)에 대해서는 후술한다.
- [0024] 최초 크로모솜 정의부(211)에서 정의된 최초 크로모솜은 데이터 완성부(202)에서 완성된 설비 데이터와 함께 모델 구축부(220)로 전달된다.

- [0025] 모델 구축부(220)는 예측 모델을 생성하고 검증하며, 데이터 분할부(221)와 예측 모델 구축부(222), 모델 적합도 생성부(223)를 포함한다.
- [0026] 데이터 분할부(221)는 데이터 완성부(202)에서 완성된 설비 데이터를 전달받아 학습 데이터와 시험 데이터로 분할한다.
- [0027] 예측 모델 구축부(222)는 모델 구축 준비부(210)로부터 입력받은 최초 크로모솜 또는 후세대 크로모솜을 이용하여 예측 모델을 구축하고, 데이터 분할부(221)의 학습 데이터를 모두 반영할 수 있도록 학습시킨다. 후세대 크로모솜에 대해서는 후세대 크로모솜 정의부(213)에서 후술한다.
- [0028] 모델 적합도 생성부(223)는 예측 모델 구축부(222)로부터 전달받은 '예측 모델'에 시험 데이터를 적용하여 모델 적합도를 생성한다.
- [0029] 모델 구축 준비부의 적합도 비교부(212)는 모델 적합도 생성부(223)로부터 생성된 모델 적합도와 미리 설정된 기준 적합도를 비교한다.
- [0030] 모델 구축 준비부의 후세대 크로모솜 정의부(213)은 적합도 비교부(212)에서 모델 적합도가 기준 적합도보다 낮은 경우 현재 정의된 크로모솜을 전세대 크로모솜으로 놓고, 이를 미리 설정된 비율에 따른 교차 및 돌연변이를 적용하여 후세대 크로모솜을 새로 정의한다. 모델 적합도가 기준 적합도보다 높거나 양 값이 같은 경우에는 현재 예측 모델을 설비 이상 예측 모델로 결정한다.
- [0031] 도 3은 본 발명에 따른 설비 이상 예측 모델의 구축 방법을 나타낸 순서도이다. 도 3의 각 단계는 도 2의 설비 이상 예측 모델 구축 시스템에 의해 수행된다.
- [0032] 단계 S300에서는, 설비 데이터를 입력받고 이를 활용하기 위해서 동기화시킨다. 이에 대한 자세한 내용은 도 4에서 후술한다.
- [0033] 단계 S400에서는, 예측 모델을 구축하기 위한 변수로 구성된 최초 크로모솜을 정의한다. 최초 크로모솜을 정의하는 것 외의 과정은 단계 S700에서 후술한다.
- [0034] 단계 S500에서는, 최초 크로모솜으로부터 예측 모델을 구축한다. 구체적으로는 예측 모델 구축에 앞서 실험 데이터를 학습 데이터와 시험 데이터로 분할하는 단계가 진행되어야 하며, 최초 크로모솜으로부터 구축된 예측 모델은 분할된 실험 데이터 중 학습 데이터에 의해 학습과정을 거친다. 학습 과정에 대한 자세한 설명은 도 5를 참조하여 후술한다.
- [0035] 단계 S600에서는, 학습과정을 거친 예측 모델에 시험 데이터를 입력하여 시험 데이터에 따른 설비 이상을 도출하고, 그로부터 모델 적합도를 산출한다.
- [0036] 단계 S700에서는, 단계 S600의 모델 적합도와 미리 설정된 기준 적합도를 비교하는 과정을 나타낸다. 기준 적합도란 예측 모델을 구축하는 반복과정을 중지시키는 데에 필요한 척도로서 허위양성(false positive)과 커버리지(coverage)로 구성된다.
- [0037] 허위양성이란, 실제로는 양성이 아닌데 양성이 나오는 정도를 비율로 나타내는 것으로서, 본 발명의 제 1실시예에서는 '90%'을 적용시키며 이 값은 실시예에 따라서 변경될 수 있다.
- [0038] 커버리지만, 예측 모델이 설비 이상이라고 판단하는 정도를 비율로 나타내는 것으로서, 본 발명의 제 1실시예에서는 '30%'을 적용시키며 이 값은 실시예에 따라서 변경될 수 있다.
- [0039] 예를 들어, 본 발명의 제 1실시예에 의할 때, 100번 설비 이상이 발생했을 경우, 설비 이상 예측 모델은 커버리지에 의해 적어도 30번이상은 설비 이상이 발생되었다고 설비 관리자에게 보고 해야하며, 설비 이상 보고에 의해 설비 관리자가 설비 이상이 있는지 직접 조사했을 때, 허위양성에 의해 실제로 적어도 27번이상은 설비 이상이 발생했어야 한다.
- [0040] 모델 적합도가 기준 적합도보다 높거나 같은 경우, 모델 결정 단계로 진행되며, 그 외의 경우는 모델 구축 준비 단계로 진행된다.
- [0041] 모델 구축 준비 단계로 진행하는 경우, 현재의 크로모솜을 전세대 크로모솜으로 정의한 후 이를 교차 및 돌연변이를 적용하여 후세대 크로모솜을 생성하는 과정을 거친다. 후세대 크로모솜은 전세대 크로모솜을 대신하여 현재의 크로모솜이 되며, 상세하게는 새로운 예측 모델을 구축할 수 있는 변수가 된다.

- [0042] 단계 S800에서는, 기준 적합도보다 높거나 같은 모델 적합도를 보이는 예측 모델을 설비 이상 예측 모델로 결정한다. 결정된 설비 이상 예측 모델은 단계 S300에서 입력되는 모든 설비 데이터에 대해서 기준 적합도 이상으로 설비 이상을 진단할 수 있는 모델이다.
- [0043] 도 4는 데이터 관리 단계에서 일어나는 과정을 상세하게 나타낸 흐름도이다.
- [0044] 단계 S310에서는, 텍스트 파일형태의 생산 데이터 파일과 트레이스 데이터 파일을 읽고 저장 및 동기화시킨다. 생산데이터란 설비에 의해 생산되는 데이터이며, 생산 데이터와 연동되어서 그 값을 가리키기 위한 데이터가 트레이스 데이터이다.
- [0045] 단계 S320에서는, 폐기되는 생산물을 수치화한 누적 버림률(accumulated loss rate)로 반응변수를 설정한다. 누적 버림률은 설비 이상이 발생했을 때의 생산물을 버리는 비율인 순간 버림률을 공정을 반복하며 누적시킨 값으로, 높을 수록 큰 반응변수가 산출되며 높은 반응변수는 곧 설비의 잦은 이상을 나타낸다. 반응 변수는 불량률, 장착 이상, 누적버림률 등으로 측정될 수 있으며 본 발명에서는 가장 경향을 알기 쉬운 누적 버림률로 반응변수를 산출한다.
- [0046] 단계 S330에서는, 완성된 동기화된 설비 데이터를 모델 구축 준비부에 전달한다. 완성된 동기화된 설비 데이터는 생산 데이터, 트레이스 데이터 및 누적버림률로 인해 정해진 반응변수를 포함한다.
- [0047] 완성된 동기화된 설비 데이터는 과거에 설비가 정상적으로 동작할 때의 데이터뿐만 아니라, 설비 이상이 발생했을 때의 데이터도 포함하게 되므로, 이 데이터를 기초로 최종적으로 구축되는 설비 이상 예측 모델에 과거에 설비 이상을 일으켰던 설비 데이터와 동일한 설비 데이터가 입력되는 경우에는 설비 데이터와 설비 이상간의 관계성과 상관없이 설비 이상을 반드시 진단할 수 있게 된다.
- [0048] 도 5는 본 발명의 실시예 중 하나인, 서포트 벡터 머신 모델을 설비 이상 예측 모델로 구축하는 과정을 나타내는 것으로, 특히 데이터 관리부를 제외한 모델 구축 준비부와 모델 생성부에서의 동작을 나타내는 흐름도이다.
- [0049] 단계 S401에서는 서포트 벡터 머신을 만들기 위한 입력 변수의 조합과 커널함수 파라미터를 선택한다. 이는 유전적 알고리즘을 진행하기 위한 최초의 단계로서, 이 두 가지 요소는 단계 S403에서 후술할 크로모솜을 구성하게 된다.
- [0050] 입력변수의 조합이란, 서포트 벡터 머신에 적용시킬 입력 벡터의 크기와 방향을 나타낸 값이다. 커널함수(kernel function)는 서포트 벡터 머신 모델이 비선형성을 가지는 두 데이터간에서도 효과적인 성능을 발휘하게끔 해주는 함수로 선형 커널(linear kernel), 다항식 커널(polynomial kernel), 방사 기저 함수 커널(RBF : Radial Basis Function kernel)등이 있으며 각각의 커널에서는 최적화를 도와주는 파라미터들이 따로 존재하고 어떠한 파라미터를 선택하는 것이 가장 좋은 지 바로 자동적으로 찾아주는 방법이 없으므로 이를 유전적 알고리즘의 반복과정을 통해 구하게 된다.
- [0051] 유전적 알고리즘에서의 선택이란, 목적으로 하는 최종해를 구하기 위해 입력하는 복수의 최초해들을 결정하는 단계로서, 룰렛 휠 선택, 토너먼트 선택, 순위 기반 선택 등 여러가지 방법이 존재한다. 최초해들이 적절하게 결정되는 경우에는 유전적 알고리즘 특유의 반복과정이 최소화되어 빠른 시간 내에 최종해를 구해낼 수 있게 된다. 예를 들어 최초해는 (15,04), (13,07), (11, 10), (09, 13) 으로 선택될 수 있다. 순서는 입력변수의 조합과 커널함수 파라미터순으로 가정한다.
- [0052] 단계 S402에서는 입력변수의 조합과 이를 이진수(binary number)로 변환시킨다. 단계 S402에서 선택된 해를 이용하면 (1111,0100), (1101,0111), (1011, 1010), (1001, 1101)이 된다.
- [0053] 단계 S403에서는 입력변수의 조합과 커널함수 파라미터를 최초 크로모솜(First Chromosome)으로 정의한다. 크로모솜은 유전적 알고리즘을 구동하기 위한 입력단위로서, 처음에 결정된 크로모솜은 진화를 거듭하여 최종해로 구성된 최종 크로모솜(Last Chromosome)을 산출하게 된다. 단계 S403에서 선택된 해를 이용하면 최초 크로모솜은 각각 11110100, 11010111, 10111010, 10011101 이 된다.
- [0054] 단계 S404에서는 최초 크로모솜과 도2의 데이터 관리부로부터 전달받은 설비 데이터가 모델 구축부로 전달된다.
- [0055] 단계 S405에서는 데이터 관리부를 거쳐 모델 구축 준비부로부터 받은 설비 데이터가 학습 데이터와 시험 데이터로 분할된다. 분할비율은 정해져 있지는 않으나, 유전적 알고리즘의 반복과정의 유효성을 위해서 학습 데이터와 시험 데이터 중 어느 한 쪽의 값이 0이어서는 안되고, 학습 데이터와 시험 데이터는 각각 설비 이상을 포함하는 설비 데이터이어야 한다. 설비 데이터를 학습 데이터와 시험 데이터로 나누는 과정은 다른 단계에 영향을 미치지

지 않으므로 모델 구축부뿐만이 아니라, 모델 구축 준비부에서 수행하는 것도 가능하다.

- [0056] 단계 S406에서는 현재의 크로모솜으로 학습 데이터의 내용을 모두 반영하는 서포트 벡터 머신 모델을 구축한다.
- [0057] 여기서 현재의 크로모솜이란 단계 S404에서 정의된 최초 크로모솜, 또는 후술할 단계 S409에 의해 정의되는 후세대 크로모솜이 될 수도 있다. 이는 단계 S409에서 후술한다.
- [0058] 학습 데이터의 내용을 모두 반영한다는 것은 구체적으로는 현재의 크로모솜으로 구축되는 서포트 벡터 머신 모델이 시험 데이터를 제외한 학습 데이터에 대해서만큼은 완전하게 설비 이상을 예측할 수 있는 모델이어야 한다는 것을 의미한다.
- [0059] 학습 데이터에는 일반 설비 동작에 관한 데이터뿐만 아니라 그에 따른 설비 이상 데이터도 포함되므로 현재의 크로모솜으로 구축되는 서포트 벡터 머신 모델이 학습 데이터에 한해서 제대로 동작하는지 판단이 가능하다. 구분을 위해서 입력변수의 조합만으로 구축된 서포트 벡터 머신 모델을 제1 예측 모델, 제1 예측 모델을 학습 데이터로 학습시켜서 학습 데이터에 한해서는 기준 적합도 이상 잘 동작하는 서포트 벡터 머신 모델을 제2 예측 모델이라고 칭할 수 있다.
- [0060] 단계 S407에서는 구축된 서포트 벡터 머신 모델에 시험 데이터를 입력하여 모델 적합도를 산출하고, 그 모델 적합도를 서포트 벡터 머신 모델과 함께 모델 구축 준비부로 다시 전달하는 과정을 나타낸다.
- [0061] 서포트 벡터 머신 모델은 학습 데이터에만 최적화되어있기 때문에 시험 데이터에 대해서 설비 이상을 완전하게 예측하리라는 보장은 없다. 그러므로 시험 데이터를 현재의 서포트 벡터 머신 모델에 입력하여 설비 이상 데이터를 산출한다. 그 다음, 서포트 벡터 머신 모델에 의해 예측된 시험 데이터에 대한 설비 이상 데이터와, 시험 데이터 내에 포함되어 있는 과거 시험 데이터에 따른 설비 이상 데이터를 비교하여 모델 적합도를 추가로 산출한다.
- [0062] 모델 적합도란 구축된 서포트 벡터 머신 모델이 설비 이상을 얼마나 잘 진단할 수 있는지 판단할 수 있는 척도로서 미리 설정된 기준 적합도와 비교를 위해 산출된다. 모델 적합도가 높다는 의미는 학습 데이터는 물론이고, 시험 데이터에 대해서도 설비 이상을 설비 관리자가 설정한 기준에 맞게 예측할 수 있다는 의미가 된다.
- [0063] 마지막으로 기준 적합도와 비교를 위해 산출된 모델 적합도와 모델 구축부에서 생성된 서포트 벡터 머신 모델은 모델 구축 준비부로 다시 전달된다.
- [0064] 단계 S408에서는 미리 설정된 기준 적합도와 단계 S407에서 산출된 모델 적합도를 비교한다.
- [0065] 단계 S409에서는 기준 적합도와 모델 적합도를 비교한 것을 기초로 하여 예측 모델 구축을 계속 할지 결정한다.
- [0066] 모델 적합도가 허위양성 90%이상과 커버리지 30%이상 중 어느 하나라도 만족시키지 못하는 경우 후세대 크로모솜을 정의한 후 단계 S406으로 진행한다.
- [0067] 후세대 크로모솜은 현재 서포트 벡터 머신 모델을 구축하는 데에 사용된 크로모솜을 전세대 크로모솜으로 놓고, 교차(crossover)와 돌연변이(mutation)를 적용시켜 산출한 크로모솜을 의미한다.
- [0068] 교차의 경우, 2점교차, 3점교차 등이 사용될 수 있고, 돌연변이의 발생확률은 최종해의 수렴에 방해가 되지 않도록 적절히 낮은 값을 설정할 수 있다.
- [0069] 예를 들어 최초 크로모솜의 한 쌍이 단계 S403에서 예로 든 11110100, 11010111 이고 2점교차가 각각 2~3번째, 7~8번째자리에 일어났다면 후세대 크로모솜은 11010110 이 된다. 여기서 돌연변이가 8번째자리에 일어나면 후세대 크로모솜으로 확정되는 것은 11010111 이 된다.
- [0070] 확정된 후세대 크로모솜은 현재의 크로모솜이 되어 단계 S406에서 새로운 서포트 벡터 머신 모델을 구축하고 기준 적합도 이상의 모델 적합도를 산출할 때까지 이후 단계를 반복하게 된다. 유전적 알고리즘을 충분히 반복하며 산출된 예측 모델의 모델 적합도가 기준 적합도에 현저히 못 미치는 경우에는 단계 S406가 아닌 단계 S405로 돌아가 설비 데이터의 분할비율이나 분할구성을 달리 할 수 있다.
- [0071] 모델 적합도가 허위양성 90%이상과 커버리지 30%이상을 모두 만족시키는 경우, 현재의 서포트 벡터 머신 모델을 설비 이상 예측 모델로 결정한다. 단계 S406에서 상술한 개념에 의하면 제2 예측 모델의 모델 적합도가 기준 적합도보다 높은 경우 그 제2 예측 모델은 설비 이상 예측 모델로 결정된다.
- [0072] 본 발명에 의해 구축된 모델을 이용하면 설비 데이터(설비의 생산정보, 이벤트 정보 또는 센서정보 등)와 설비 이상간에 비선형관계를 보이더라도 실시간으로 설비 데이터로부터 설비 이상을 높은 확률로 예측할 수 있으므로

해당 설비를 관리하는 데에 도움을 줄 수 있으며, 특히 누적버림률의 경향성이 명확한 표면 장착 기술(SMT : Surface Mounting Technology) 설비의 장착 이상을 실시간으로 진단하는 데에 우수성을 갖는다.

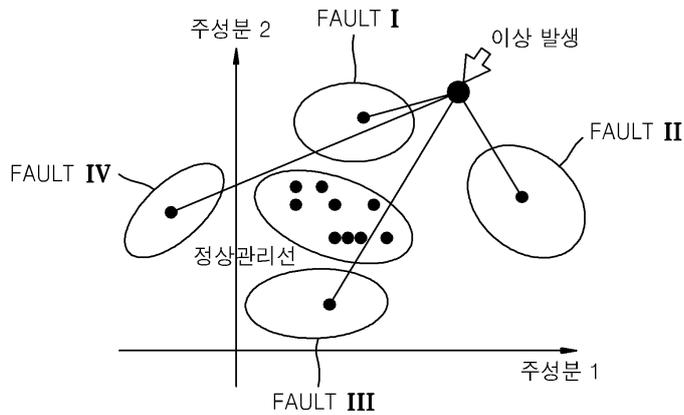
[0073] 본 발명은 상술한 실시형태 및 첨부된 도면에 의해 한정되지 아니므로 상술된 서포트 벡터 머신 외의 다른 기계 학습 알고리즘에 의해서도 본 발명은 구현될 수 있다. 첨부된 청구범위에 의해 권리범위를 한정하고자 하며, 청구범위에 기재된 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 다양한 형태의 치환, 변형 및 변경할 수 있다는 것은 당 기술분야의 통상의 지식을 가진 자에게 자명할 것이다.

부호의 설명

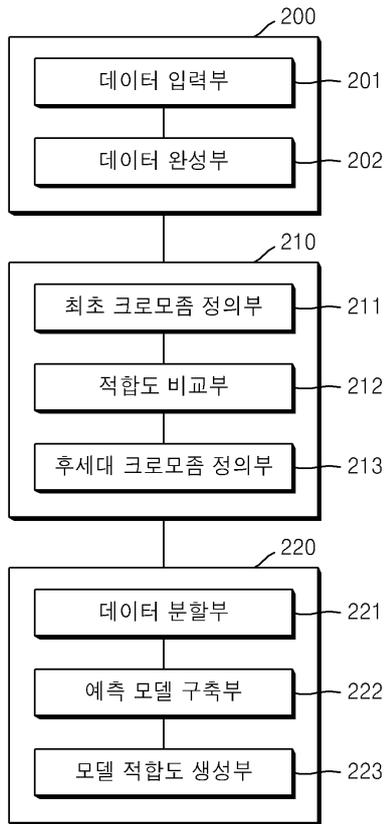
- [0074] 200 : 데이터 관리부
- 201 : 데이터 입력부
- 202 : 데이터 완성부
- 210 : 모델 구축 준비부
- 211 : 최초 크로모즘 정의부
- 212 : 적합도 비교부
- 213 : 후세대 크로모즘 정의부
- 220 : 모델 구축부
- 221 : 데이터 분할부
- 222 : 예측 모델 구축부
- 223 : 모델 적합도 생성부

도면

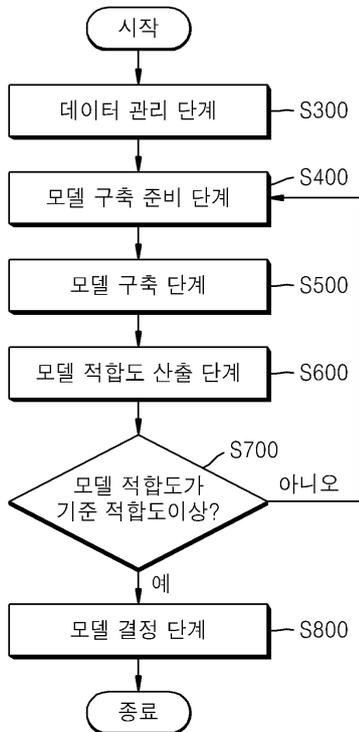
도면1



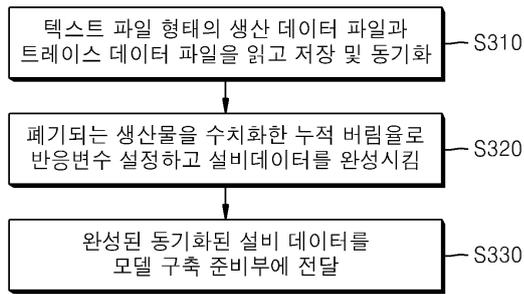
도면2



도면3



도면4



도면5

