



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0053692
(43) 공개일자 2017년05월16일

(51) 국제특허분류(Int. Cl.)
G06F 17/18 (2006.01) G06K 9/00 (2006.01)
(52) CPC특허분류
G06F 17/18 (2013.01)
G06K 9/00 (2013.01)
(21) 출원번호 10-2017-7009536
(22) 출원일자(국제) 2015년03월04일
심사청구일자 없음
(85) 번역문제출일자 2017년04월07일
(86) 국제출원번호 PCT/US2015/018698
(87) 국제공개번호 WO 2016/039805
국제공개일자 2016년03월17일
(30) 우선권주장
62/049,558 2014년09월12일 미국(US)

(71) 출원인
지이 인텔리전트 플랫폼스 인코포레이티드
미국 버지니아주 22911 살로츠빌 오스틴 드라이브
2500
(72) 발명자
헤르조그 제임스 피.
미국 일리노이즈주 60532 라일 스위트 300 워런빌
로드 901
(74) 대리인
김태홍, 김진희

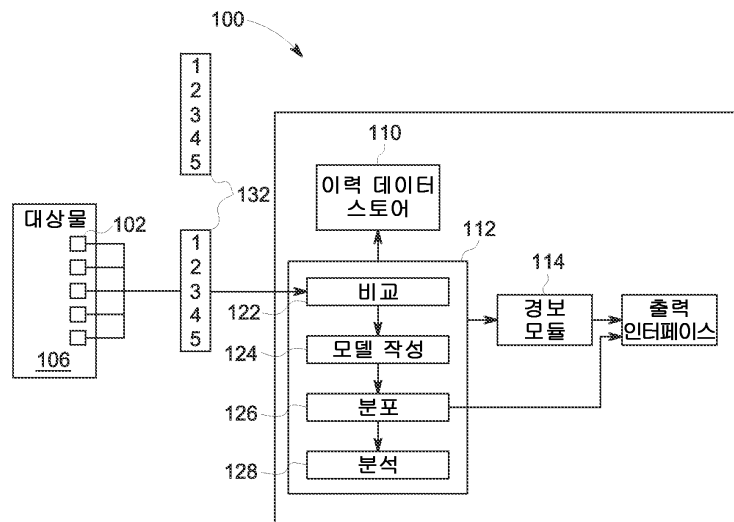
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 커널 회귀 모델의 앙상블을 위한 장치 및 방법

(57) 요약

엔티티 또는 프로세스와 연관된 물리적 파라미터를 나타내는 정보가 감지된다. 감지된 정보는 현재 패턴에 또는 패턴의 현재 시퀀스에 수집된다. 현재 패턴 또는 패턴의 현재 시퀀스는 베스트 매칭의 모집단을 얻기 위해 이력 데이터와 비교된다. 베스트 매칭의 모집단에 기초하여 복수의 커널 회귀 모델이 작성된다. 복수의 커널 회귀 모델을 이용하여 관심 센서에 대해 추정치의 적어도 하나의 분포가 생성된다. 적어도 하나의 추정치 분포의 중심의 척도 및 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 얻기 위해 관심 센서에 대한 추정치의 적어도 하나의 분포가 분석된다.

대표도 - 도1



명세서

청구범위

청구항 1

엔티티 또는 프로세스의 현재 또는 미래의 거동(behavior)을 추정하는 방법에 있어서,
 상기 엔티티 또는 프로세스와 연관된 물리적 파라미터를 나타내는 정보를 감지하는 단계와,
 감지된 정보를 현재 패턴에 또는 패턴의 현재 시퀀스에 수집하는 단계와,
 베스트 매칭의 모집단(population of best match)을 얻기 위해 상기 현재 패턴 또는 패턴의 현재 시퀀스를 이력 데이터와 비교하는 단계와,
 상기 베스트 매칭의 모집단에 기초하여 복수의 커널 회귀 모델을 작성하는 단계와,
 상기 복수의 커널 회귀 모델을 이용하여 관심 센서에 대한 추정치의 적어도 하나의 분포를 생성하는 단계와,
 관심 센서에 대한 상기 추정치의 적어도 하나의 분포를 분석하여 적어도 하나의 추정치 분포의 중심의 척도(measure) 및 상기 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 취득하는 단계를 포함하는, 엔티티 또는 프로세스의 현재 또는 미래의 거동을 추정하는 방법.

청구항 2

제1항에 있어서, 상기 작성하는 단계는 현재의 단일 시점(時點)에서 상기 복수의 커널 회귀 모델을 작성하는 단계를 포함하는 것인 방법.

청구항 3

제1항에 있어서, 상기 작성하는 단계는, 상기 현재의 단일 시점에서 끝나는 관련된 시점(時點)들의 시간적 시퀀스에 대해 상기 복수의 커널 회귀 모델을 작성하는 단계를 포함하는 것인 방법.

청구항 4

제1항에 있어서, 상기 적어도 하나의 추정치 분포의 상기 중심의 척도는 평균(average)을 포함하는 것인 방법.

청구항 5

제1항에 있어서, 상기 적어도 하나의 추정치 분포의 상기 중심의 척도는 중앙값(median)을 포함하는 것인 방법.

청구항 6

제1항에 있어서, 상기 추정치 분포 폭의 척도는 표준 편차를 포함하는 것인 방법.

청구항 7

제1항에 있어서, 미리 정해진 기준에 기초하여 상기 복수의 커널 회귀 모델 중 적어도 하나를 선택적으로 제거하는 단계를 더 포함하는 방법.

청구항 8

추정치를 취득하는 장치에 있어서,
 입력부와 출력부를 구비한 인터페이스로서, 상기 입력부는 엔티티 또는 프로세스와 연관된 물리적 파라미터를 나타내는 감지된 정보를 수신하도록 구성되며, 상기 감지된 정보는 현재 패턴에 또는 패턴의 현재 시퀀스에 수집되는 것인, 상기 인터페이스와,
 상기 인터페이스에 연결된 프로세서를 포함하고,

상기 프로세서는, 베스트 매칭의 모집단을 취득하기 위해 상기 현재의 패턴 또는 패턴의 현재 시퀀스를 이력 데이터와 비교하도록 구성되고, 상기 프로세서는, 상기 베스트 매칭의 모집단에 기초하여 복수의 커널 회귀 모델을 작성하고, 상기 복수의 커널 회귀 모델을 이용하여 관심 센서에 대한 추정치의 적어도 하나의 분포를 생성하도록 구성되며, 상기 프로세서는 또한, 관심 센서에 대한 상기 추정치의 적어도 하나의 분포를 분석하여 적어도 하나의 추정치 분포의 중심의 척도와 상기 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 취득하고, 상기 적어도 하나의 추정치 분포의 중심의 척도와 상기 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 상기 출력부에 제공하도록 구성되는 것인 추정치를 취득하는 장치.

청구항 9

제8항에 있어서, 상기 복수의 커널 회귀 모델은 현재의 단일 시점에서 작성되는 것인 장치.

청구항 10

제8항에 있어서, 상기 복수의 커널 회귀 모델은 상기 현재의 단일 시점에서 끝나는 관련된 시점들의 시간적 시퀀스에 대해 작성되는 것인 장치.

청구항 11

제8항에 있어서, 상기 적어도 하나의 추정치 분포의 상기 중심의 척도는 평균을 포함하는 것인 장치.

청구항 12

제8항에 있어서, 상기 적어도 하나의 추정치 분포의 상기 중심의 척도는 중앙값을 포함하는 것인 장치.

청구항 13

제8항에 있어서, 상기 추정치 분포 폭의 척도는 표준 편차를 포함하는 것인 장치.

청구항 14

제8항에 있어서, 상기 프로세서는, 미리 정해진 기준에 기초하여 상기 복수의 커널 회귀 모델 중 적어도 하나를 선택적으로 제거하도록 구성되는 것인 장치.

발명의 설명

기술 분야

[0001] <관련 출원과의 교차 참조>

[0002] 본 출원은 35 U.S.C. § 119 (e)에 따라 2014년 9월 12에 출원한 발명의 명칭이 APPARATUS AND METHOD FOR ENSEMBLES OF KERNEL REGRESSION MODELS인 미국 가출원 번호 62/049558에 대해 우선권을 주장하며, 이 우선권의 내용은 그 전체가 참조문헌으로 포함된다.

[0003] <발명의 분야>

[0004] 본 출원은 모델링에 관한 것이며, 더 구체적으로는 모델링에 기초하여 파라미터의 거동(behavior)의 추정치를 획득하는 것에 관한 것이다.

배경 기술

[0005] 커널 회귀법(kernel regression)은 데이터세트 내의 값들 사이에서 비선형 함수 또는 관계를 결정하는데 사용되는 모델링의 한 형태로서, 머신 또는 시스템을 모니터링하여 머신 또는 시스템의 상태를 결정하는데 사용된다. SSM(Sequential Similarity Based Modeling)의 경우, 다수의 센서 신호는 모니터링되고 있는 머신, 시스템, 또는 기타 대상물의 물리적 상관 파라미터를 측정하여 센서 데이터를 제공한다. 파라미터 데이터는 신호로부터의 실제 또는 현재 값, 또는 센서 신호에 기초하거나 하지 않는 기타 산출 데이터를 포함할 수 있다. 그런 다음 파라미터 데이터는 경험적 모델에 의해 프로세싱되어 해당 값의 추정치를 제공한다. 그리고, 모니터링되고 있는 시스템에 오류(fault)가 존재하는지의 여부를 판정하기 위해 추정치를 실제 또는 현재 값과 비교한다.

[0006] 보다 구체적으로, 모델은 알려진 동작 상태를 나타내는 센서 값의 선택된 이력 패턴의 기준 라이브러리를 사용

하여 추정치를 생성한다. 이들 패턴은 벡터, 스냅샷 또는 관찰치라고도 칭해지며, 시간의 한 시점(on instant in time)에 모니터링되고 있는 머신의 상태를 나타내는, 다수의 센서로부터의 값 또는 기타 입력 데이터를 포함한다. 기준 라이브러리로부터의 기준 벡터의 경우, 벡터는 대개 모니터링되고 있는 머신의 정상 동작을 나타낸다. 모델은 현재 시간으로부터의 벡터를 기준 라이브러리의 알려진 상태로부터의 다수의 선택된 학습 벡터와 비교하여 시스템의 현재 상태를 추정한다. 일반적으로 말하자면, 현재 벡터는 기준 라이브러리로부터의 선택된 벡터로 이루어진 행렬과 비교되어 가중 벡터를 형성한다. 추가 단계에 있어서, 가중 벡터는 추정 값의 벡터를 산출하기 위한 행렬과 곱해진다. 그런 다음, 추정 벡터는 현재 벡터와 비교된다. 벡터의 추정 값과 실제 값이 충분히 유사하지 않다면, 이것은 모니터링되고 있는 대상물에 오류가 존재하는 것을 나타낼 수 있다.

[0007] 다른 형태의 커널 회귀 모델링은 VBM(Variable Similarity Based Modeling)이다. VBM에서는, 기준 데이터 관찰치가 머신, 프로세스 또는 시스템을 나타내는 센서 또는 측정치로부터 먼저 취득된다. 그런 다음, 그 나타내는 데이터와, 동일한 센서 또는 측정치로부터의 현재의 관찰치와의 조합으로부터 모델이 산출된다. 모델은 모델링된 시스템의 각각의 새로운 관찰치에 의해 재산출된다. 모델의 출력은 모델링된 시스템의 상태를 특성화하는 적어도 하나의 센서, 측정치, 또는 기타 분류(classification) 또는 자격(qualification) 파라미터의 추정치이다.

[0008] 전술한 접근법을 이용하여 추정치를 얻을 수도 있지만, 이러한 방식으로 추정치를 얻는 데에는 몇몇 제한이 있다. 일부 업계에서는, 회귀 모델을 사용하여, 중요한 시구간에 측정되지 않거나 전혀 측정할 수 없는 핵심 센서 또는 동작 파라미터의 응답을 추정함에 있어서, 미래 응답이 추정되고 있기 때문에, 문제가 있다. 신뢰 한계(confidence bound)의 정확한 산출은, 추정치 및 연관된 신뢰 한계가 주요 파라미터에 대해 이용 가능한 유일한 데이터이기 때문에 이러한 문제에 특히 유용하다.

[0009] 전술한 업계의 문제들 중 일례는 펌프를 이용한(pump-assisted) 석유 및 가스 추출에 관한 것이다. 유정 내에 그리고 전기 잠수정 펌프 상에 있는 다운홀 센서가 용기(reservoir) 온도, 용기 압력, 및 펌프 속도 등의 파라미터의 연속 측정을 제공하지만, 추출된 석유 및 가스의 양을 결정하는 데에 이용되는 주요 유정 성능 파라미터는 없다. 체적 유량 및 워터-컷(water-cut)(즉, 유정으로부터 생산된 총 액체의 부피 대비 생산된 물의 비율) 등의 주요 성능 파라미터는 유정 테스트 시에 불규칙한 간격으로 측정된다. 결과적으로, 현재의 접근법은 이러한 유형의 추정치를 취득할 때 적절하거나 허용되는 작업을 행하지 않는다.

[0010] 이러한 문제들 때문에 이전의 접근법에 대해 일부 일반적 사용자 불만족이 생겼다.

발명의 내용

[0011] 본 접근법은 모니터링되고 있는 대상물 또는 프로세스로부터 수신되는 센서 데이터의 각각의 관찰 벡터마다 커널 회귀 모델의 앙상블(패밀리)을 생성한다. 앙상블 내의 모델은 현재의 조건과 유사하지만 서로 독립적인 데이터로부터 작성된다. 각각의 모델은 각각의 모델 변수마다 추정 벡터를 생성한다. 각 변수마다 생성된 추정치의 분포로부터 통계치가 산출된다. 일 양태에 있어서, 추정치 분포의 평균이 계산되고, 이것은 임의의 단일 모델에 의해 생성되는 것보다 현재 상태의 더 강력한 추정치를 제공한다. 다른 양태에 있어서, 분포의 중앙값(median)이 계산된다. 독립적인 모델의 모집단(population)이 센서 및 프로세스 에러와 상관되기 때문에, 추정치 분포의 폭의 척도(measure)(예컨대, 표준 편차)는 현재의 관찰 벡터에 대한 모델 추정치의 불확정성의 지표를 제공한다.

[0012] 이들 실시형태의 다수에 있어서, 엔티티 또는 프로세스와 연관된 물리적 파라미터를 나타내는 정보가 감지된다. 감지된 정보는 현재 패턴에 또는 패턴의 현재 시퀀스에 수집된다. 현재 패턴 또는 패턴의 현재 시퀀스는 베스트 매칭의 모집단을 얻기 위해 이력 데이터와 비교된다. 베스트 매칭의 모집단에 기초하여 복수의 커널 회귀 모델이 작성된다. 복수의 커널 회귀 모델을 이용하여 적어도 하나의 관심 센서에 대해 추정치의 적어도 하나의 분포가 생성된다. 관심 센서 각각에 대해 추정치 분포의 중심의 척도와 추정치 분포의 폭의 척도를 얻기 위해 하나 이상의 관심 센서에 대해 추정치의 분포가 분석된다.

[0013] 일부 양태에 있어서, 복수의 커널 회귀 모델을 작성하는 단계는 현재의 단일 시점(時點)에서 복수의 커널 회귀 모델을 작성하는 단계를 포함한다. 다른 양태에 있어서, 복수의 커널 회귀 모델을 작성하는 단계는 현재의 단일 시점에서 끝나는 관련된 시점(時點)들의 시간적 시퀀스에 대해 복수의 커널 회귀 모델을 작성하는 단계를 포함한다.

[0014] 일부 예에 있어서, 추정치 분포의 중심의 척도는 평균(average)을 포함한다. 다른 예에 있어서, 추정치 분포의 중심의 척도는 중앙값(median)을 포함한다. 다른 양태에 있어서, 추정치 분포 폭의 척도는 표준 편차를 포함한다.

다. 일부 다른 예에 있어서, 복수의 모델 중 적어도 하나는 미리 정해진 기준에 기초하여 선택적으로 제거된다.

[0015] 이들 실시형태 중 다른 것에 있어서, 추정치를 취득하는 장치는 인터페이스와 프로세서를 포함한다. 인터페이스는 입력부와 출력부를 포함하며, 입력부는 엔티티 또는 프로세스와 연관된 물리적 파라미터를 나타내는 감지된 정보를 수신하도록 구성된다. 감지된 정보는 현재 패턴에 또는 패턴의 현재 시퀀스에 수집된다.

[0016] 프로세서는 인터페이스에 연결된다. 프로세서는 베스트 매칭의 모집단을 얻기 위해 현재 패턴 또는 패턴의 현재 시퀀스를 입력 데이터와 비교하도록 구성된다. 프로세서는, 베스트 매칭의 모집단에 기초하여 복수의 커널 회귀 모델을 작성하고 복수의 커널 회귀 모델을 이용하여 관심 센서에 대한 추정치의 적어도 하나의 분포를 생성하도록 구성된다. 프로세서는 또한, 관심 센서에 대한 추정치의 적어도 하나의 분포를 분석하여 적어도 하나의 추정치 분포의 중심의 척도 및 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 취득하도록 구성된다. 프로세서는 적어도 하나의 추정치 분포의 중심의 척도 및 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 출력부에 제공한다.

도면의 간단한 설명

[0017] 개시내용의 보다 완전한 이해를 위해서는, 이하의 상세한 설명과 첨부하는 도면을 참조해야 한다.

도 1은 본 발명의 다양한 실시형태에 따라 추정치를 취득하는 시스템의 블록도를 포함한다.

도 2는 본 발명의 다양한 실시형태에 따라 추정된 값의 상이한 통계적 양상을 나타내는 그래프를 포함한다.

도 3은 본 발명의 다양한 실시형태에 따라 추정치를 취득하는 접근법의 흐름도를 포함한다.

도 4는 본 발명의 다양한 실시형태에 따라 추정치를 취득하는 장치의 블록도를 포함한다.

숙련된 기술자는 도면의 구성요소가 단순하고 명료하게 도시되어 있음을 이해할 것이다. 또한, 소정의 동작 및/또는 단계가 특정 발생 순서로 기재 또는 묘사 될 수 있다는 것이 이해되겠지만, 당업자에게는 시퀀스에 대한 그러한 특이성이 실제로 요구되지 않는다는 것도 이해될 것이다. 또한, 본 명세서에서 사용된 용어 및 표현은 특정 의미가 달리 명시된 경우를 제외하고는 해당하는 각각의 조사 및 연구 분야에서의 그러한 용어 및 표현에 부합하는 통상적인 의미를 갖는다는 것이 이해될 것이다.

발명을 실시하기 위한 구체적인 내용

[0018] 본 접근법은 랜덤 포레스트(random forest) 및 그레디언트 부스팅(gradient boosting) 모델과 같은 확률적 모델링 방법의 특이성인 앙상블 학습(ensemble learning) 및 랜덤화된 피쳐 선택 속성(randomized feature selection attribute)을 이용한다. 그러나, 결정 트리과 같은 열등 학습자(weak learner)를 이용하는 이들 전통적인 앙상블 학습 알고리즘과 달리, 본 접근법은 국소화된 커널 회귀 모델의 비교적 강력한 학습 알고리즘을 이용한다.

[0019] 커널 회귀 모델링 알고리즘의 두 가지 형태는 국소화된 학습 알고리즘을 이용하는데, 이들 모델링 알고리즘 둘 다는 본 발명의 접근법에 따라 이용될 수 있다. VBM(Variable Similarity Based Modeling)이라고도 알려진, 이들 모델링 알고리즘의 제1 형태의 일례는, 그 전체가 참조문헌으로 포함되는 미국 특허 7,403,869에 기술되어 있다. SSM(Sequential Similarity Based Modeling)이라고도 알려진 커널 회귀 알고리즘의 제2 형태의 일례는, 그 전체가 참조문헌으로 포함되는 미국 특허 8,602,853에 기술되어 있다.

[0020] 본 접근법에 의해 이용되는 국소화된 학습 알고리즘에서는, 모니터링되는 시스템의 현재 상태가 훨씬 더 많은 학습 상태의 기준 어레이의 상태와 비교된다. 현재 상태와 기준 어레이의 각각의 상태 간의 패턴 중첩(pattern overlap)의 스코어치를 제공하기 위해 유사성 연산자 또는 기타 패턴 매칭 함수가 적용된다. 최고 스코어를 가진 기준 상태의, 예컨대 10개로 이루어진, 소 세트(small set)가 트레이닝 행렬에 수집되어 모델을 작성한다. 이 모델은 현재 상태의 추정치를 생성하는데 이용된다.

[0021] VBM 알고리즘의 상황에서는, 상태(state)가 관찰 벡터이지만, SEM의 상황에서는, 상태가 시간과 관련된(temporally-related) 관찰 벡터의 시퀀스이다. 본 명세서의 대부분에 있어서, 대부분의 설명은 VBM 알고리즘을 이용하는 본 접근법의 적용에 관한 것이다. 그러나 보편성의 손실 없이, 본 접근법은 SSM 알고리즘에도 동등하게 적용되고 SSM 알고리즘을 이용할 수 있다는 것을 이해해야 한다.

[0022] 기준 어레이 내의 벡터 수가 시스템의 고유의 동작 상태의 수보다 많은 경향이 있기 때문에, 현재의 관찰 벡터와 굿 매칭(good match)하는 기준 벡터의 작은 부분만이 선택된다. 또한, 최고의 패턴 매칭을 생성하는 기준 벡

터는 관찰 벡터의 랜덤한 변동(random fluctuation)과 일치하는 랜덤한 변동을 갖는 경향이 있다. 이러한 합성 신호에서의 랜덤 요소들의 정렬(alignment)은 모델이 데이터의 노이즈 성분과 오버피팅(overfit)하는 경향을 증가시킨다.

- [0023] 본 명세서에서 설명하는 앙상블 커널 회귀 모델 기반 접근법은, 관찰 벡터와 꺾 매칭하는 더 많은 모집단의 기준 벡터 중에서 트레이닝 벡터를 랜덤하게 선택함으로써, 오버피팅하는 모델을 생성하는 국소화된 학습 알고리즘의 경향을 방해한다. 회귀 모델을 작성하기 위한 기준 벡터의 랜덤한 선택은 다수 회, 예컨대 50회 수행된다.
- [0024] 각각의 회귀 모델은 추정 벡터를 생성한다. 커널 회귀 모델의 앙상블에 의해 생성된 추정 벡터의 집합은 구성 벡터 중 어느 것보다 잡음에 의해 덜 착색된 추정 벡터를 생성하도록 평균화된다. 모델의 앙상블의 정확성은 표준 편차 또는 분포의 5번째와 95번째 백분위 사이의 차이 등의 추정 벡터의 분포의 변화의 척도로서 제공된다. 이들 통계치가 모델의 각 변수마다 산출된다.
- [0025] 트레이닝 벡터가 랜덤하게 선택되기 때문에, 앙상블 모델이 불완전하게 수행할 가능성이 있다. 일부 양태에 있어서, 임의의 불완전하게 수행되는 앙상블 모델을 제거하기 위해 가지치기 알고리즘(pruning algorithm)이 이용된다. 일례에 있어서, 가지치기 알고리즘은 전역 유사성(global similarity)이라 불리는 통계치를 이용하며, 이것은 그 전체가 참조문헌으로 포함되는 미국 특허 6,859,739에 기술되어 있다. 다른 유형의 가지치기 알고리즘도 존재한다. 일반적으로, 이들 알고리즘은 모델 품질 또는 피팅 양호성(goodness of fit)의 통계적 척도를 제공한다. 이러한 통계적 척도는 평균 제곱근 에러 및 결정 계수(R 제곱 통계치라고도 함) 등의 척도를 포함한다. 가지치기 알고리즘은 각각의 앙상블 모델(즉, 추정 벡터)의 출력에 모델 품질 척도를 적용하고, 일부 미리 정해진 임계값보다 품질이 낮은 임의의 앙상블 모델을 제거한다.
- [0026] 모델 추정치가 관련 모델의 패밀리리의 평균 응답으로부터 도출되므로, 앙상블 커널 회귀 모델은 관찰 벡터에 대한 단일 추정치를 생성하는 표준 커널 회귀 모델보다 더 강력한 시스템 응답 추정치를 제공하는데, 단일 모델에 영향을 미치는 프로세스와 센서 잡음이 앙상블을 평균화함으로써 감소하기 때문이다. 그러나 더 큰 장점은 모델 출력의 앙상블에 걸친 변동이 전체 모델 응답의 직접적인 신뢰도 척도라는 것이다. 앙상블 커널 회귀 모델은 모든 모델 변수의 응답 추정치뿐만 아니라, 개별 추정치에 대한 상위 및 하위 신뢰 한계도 제공할 수 있다.
- [0027] 도 1을 참조하면, 시간 도메인 정보를 통합하는 SSM 시스템 또는 VBM 시스템일 수 있는 추정 시스템(100)은 하나 이상의 모듈의 형태로 컴퓨터 프로그램 내에 포함되어, 하나 이상의 컴퓨터 상에서 그리고/또는 하나 이상의 프로세서에 의해 실행될 수 있다.
- [0028] 컴퓨터 또는 프로세서는 영구적이든 일시적이든 센서 데이터 및/또는 컴퓨터 프로그램을 보유하기 위해 내부 또는 외부의 하나 이상의 메모리 저장 디바이스를 가질 수 있다. 일 형태에 있어서, 독립형 컴퓨터(standalone computer)는, 기기가 탑재된 머신(instrumented machine), 프로세스 또는 생체를 비롯한 기타 대상물 상에 있는 센서로부터 센서 데이터를 수신하여, 파라미터(온도, 압력 등)를 측정하기 위한 전용 프로그램을 실행한다. 모니터링되고 있는 대상물은 특히 제한되지는 않지만, 몇가지 예를 들자면, 풍력 발전단지(wind farm) 내의 하나 이상의 풍력 터빈, 해저 유정과 관련된 장비, 산업용 플랜트 내의 하나 이상의 머신, 하나 이상의 차량, 또는 차량 상의 제트 엔진과 같은 특정 머신일 수 있다. 센서 데이터는 유선 또는 무선으로 예컨대 컴퓨터 네트워크 또는 인터넷을 통해 데이터 수집을 수행하는 컴퓨터 또는 데이터베이스에 전송될 수 있다. 하나 이상의 프로세서를 구비한 하나의 컴퓨터 또는 프로세서는 모든 모듈에 대해 모든 모니터링 태스크를 수행할 수 있거나, 각 태스크 또는 모듈은 그 자체 컴퓨터 또는 그 모듈을 수행하는 프로세서를 가질 수 있다. 따라서, 프로세싱은 단일 위치에서 일어날 수도 있고 또는 유선 또는 무선 네트워크에 의해 모두 접속되는 다수의 상이한 위치에서 일어날 수도 있다는 것이 이해될 것이다.
- [0029] 시스템(100)은 전술한 바와 같이 모니터링되고 있는 대상물(106) 상의 센서(102)로부터 데이터 또는 신호를 수신한다. 이 신호는 시스템(100)에 의해 사용할 하나 이상의 입력 벡터(132)로 배열된다. 여기서, 입력, 실제, 및 현재라는 용어는 상호 교환적으로 사용되며, 벡터, 스냅샷 및 관찰치라는 용어는 상호 교환적으로 사용된다. 입력 벡터(또는 예컨대 실제 스냅샷)는 시간의 한 순간에 모니터링되고 있는 머신의 동작 상태를 나타낸다. 일 예에서는, 하나의 입력 벡터가 수신된다(VBM). 다른 예에서는 시간과 관련된 벡터의 시퀀스가 수신된다(SSM). 일 예에서는, 여러 센서 값이 매우 빈번하게 얻어지지만, 다른 센서 값은 드물게 얻어진다. 다시 말해, 현재의 시점(point in time)에서는 일부 센서 값이 명확하게 알려져 있지만 다른 센서 값은 알려져 있지 않다.
- [0030] 사용자 편에서는 현재의 시점에서 하나 이상의 관심 센서로부터 드물게 얻어지는(미지의) 센서 값의 추정치를 얻는 것이 바람직하다. 또한, 미래의 시점에서 하나 이상의 관심 센서로부터 드물게 얻어지는(미지의) 센서 값

에 대한 추정치를 얻는 것도 바람직할 수 있다. 이들 결과 둘 다에 대해, 추정된 값의 통계적 불확정성을 아는 것이 바람직하다. 본 명세서에서 설명하는 접근법을 이용하여, 이 정보가 확인되어 출력 인터페이스(116)에서 사용자에게 제시될 수 있다.

[0031] 입력 벡터(132)는 센서 데이터(또는 미처리 데이터)에 기초하여 산출되거나 산출되지 않을 수도 있는 계산된 데이터를 포함할 수 있다. 이것은 예컨대, 평균 압력 또는 압력의 변화, 온도, 풍속, 유속 및 기타 유형의 계산된 파라미터를 포함할 수 있다. 입력 벡터(132)는 또한 대상물(106) 상의 센서가 나타내지 않는 다른 변수를 나타내는 값도 가질 수 있다. 이것은 예컨대 센서 데이터가 수신된 년도의 그 날의 평균 주변 온도 동일 수 있다.

[0032] 시스템은 입력 데이터 스토어(110), 추정 모듈(112), 경보 모듈(114), 및 출력 인터페이스(116)를 포함한다. 추정 모듈(112)은 비교 모듈(122), 모델 작성 모듈(124), 분포 모듈(126), 및 분석 모듈(128)을 포함한다. 구성요소들 중 어떤 것은 하드웨어 및/또는 컴퓨터 소프트웨어의 임의의 조합을 이용해서 구현될 수 있음을 알 것이다. 예를 들어, 구성요소들 중 어떤 것은 프로세싱 디바이스 상에서 실행되는 컴퓨터 명령어를 이용하여 구현될 수도 있다.

[0033] 동작 시에, 데이터가 추정 모듈(112)에 의해 수신된다. 추정 모듈은 추정치와, 그 추정치에 대한 정확성 범위를 제공한다. 추정치와 정확성 범위는 현재의 시점에 관한 것이거나(VBM 접근법이 사용되는 경우), 하나 이상의 미래 시점에 관한 것일 수 있다(SSM 접근법이 사용되는 경우). 경보 모듈(114)은 소정의 미리 정해진 기준이 충족할 때에 사용자에게 경보를 보낼 수 있다. 추정치(그리고 분포/추정치의 불확정성)와 함께 경보가 출력 인터페이스(116)에 표시될 수 있다. 출력 인터페이스(116)는 임의 유형의 디바이스(예컨대, 컴퓨터, 태블릿, 셀룰러 폰, 디스플레이) 상의 임의 유형의 인터페이스(예컨대, 디스플레이 스크린, 터치 스크린)일 수 있다.

[0034] 이제 추정 모듈(112)의 특정 동작 및 구조를 살펴보면, 전술한 바와 같이 일 양태에 있어서, 그것의 기능을 수행하기 위해 4개의 모듈(122, 124, 126, 128)이 이용된다. 모듈(122, 124, 126, 128)은 하드웨어와 소프트웨어의 임의의 조합으로 구현될 수 있음이 이해될 것이다. 일례에 있어서, 모듈(122, 124, 126, 128)은 마이크로프로세서 등의 프로세싱 디바이스 상에서 수행되는 컴퓨터 명령어를 이용하여 구현된다.

[0035] 비교 모듈(122)은 (수신된 입력 벡터로부터 취득된) 현재 패턴 또는 패턴의 현재 시퀀스를 입력 데이터 스토어로부터의 입력 데이터와 비교하여 베스트 매칭의 모집단을 취득한다. 베스트 매칭은 미리 정해진 기준을 충족하는 것일 수 있다. 예를 들어, 소정의 임계치 이상의 유사도 값을 갖는 벡터가 선택될 수 있다.

[0036] 모델 작성 모듈(124)은 베스트 매칭의 모집단에 기초하여 복수의 커널 회귀 모델을 작성한다. 다음의 수식과 설명은 유사성 기반 모델(SBM)에 대한 것이다. SBM은 커널 회귀 모델링의 한 형태이다. 다른 형태의 커널 회귀 모델도 이용될 수 있음이 이해될 것이다.

[0037] 여기에 언급하는 모델은 데이터 구조로서 구현되거나 저장될 수 있는 수학적 관계를 나타낸다. 추정치는 이러한 모델로부터 만들어지고, 추정치는, 다음의 수식에 따라, 데이터의 출처와 독립적으로 만들어지는데, 추정치는 유사도 연산자로부터 생성된 "가중치"의 합으로 나누어서 정규화된다.

$$\mathbf{x}_{est} = \frac{\mathbf{D} \cdot (\mathbf{D}^T \otimes \mathbf{D})^{-1} \cdot (\mathbf{D}^T \otimes \mathbf{x}_{new})}{\sum ((\mathbf{D}^T \otimes \mathbf{D})^{-1} \cdot (\mathbf{D}^T \otimes \mathbf{x}_{new}))}$$

[0038] (1)

[0039] 유사성 기반 모델링의 추론 형태에서는, 추론된 파라미터 벡터 $\mathbf{y}_{est}$ 가 다음과 같이, 학습된 관찰치와 입력으로부터 추정된다.

$$\mathbf{y}_{est} = \mathbf{D}_{out} \cdot (\mathbf{D}_{in}^T \otimes \mathbf{D}_{in})^{-1} \cdot (\mathbf{D}_{in}^T \otimes \mathbf{x}_{in})$$

[0040] (2)

[0041] 여기서, $\mathbf{D}_{in}$ 는 $\mathbf{x}_{in}$ 의 실제 센서 값(또는 파라미터)과 동일한 수의 행을 갖고, $\mathbf{D}_{out}$ 는 추론된 파라미터 또는 센서를 포함한 파라미터의 총수와 같은 수의 행을 갖는다.

[0042] 한 형태에 있어서, 학습된 표본 $\mathbf{D}_a$ 의 행렬은 입력 벡터 $\mathbf{x}_{in}$ 의 센서 값에 매핑되는 행과 다양한 센서에 매핑되는 행을 둘 다 포함하는 집계 행렬로 이해될 수 있다.

$$\mathbf{D}_a = \begin{bmatrix} \mathbf{D}_{in} \\ \mathbf{D}_{out} \end{bmatrix}$$

[0043] (3)

[0044] 앞에서와 같이 가중치의 합계를 이용해, 정규화하면 다음과 같다.

$$\mathbf{y}_{est} = \frac{\mathbf{D}_{out} \cdot (\mathbf{D}_{in}^T \otimes \mathbf{D}_{in})^{-1} \cdot (\mathbf{D}_{in}^T \otimes \mathbf{x}_{in})}{\sum ((\mathbf{D}_{in}^T \otimes \mathbf{D}_{in})^{-1} \cdot (\mathbf{D}_{in}^T \otimes \mathbf{x}_{in}))} \quad (4)$$

[0045]

[0046] $\mathbf{D}_{out}$ 을 학습된 표본 $\mathbf{D}_a$ 의 전체 행렬로 대체함으로써, 유사성 기반 모델링이 입력 센서(자동 연관형) 및 유추 센서(추론형)에 대한 추정치를 동시에 산출할 수 있다는 것을 알아야 한다.

$$\begin{bmatrix} x_{est} \\ y_{est} \end{bmatrix} = \frac{\mathbf{D}_a \cdot (\mathbf{D}_{in}^T \otimes \mathbf{D}_{in})^{-1} \cdot (\mathbf{D}_{in}^T \otimes \mathbf{x}_{in})}{\sum ((\mathbf{D}_{in}^T \otimes \mathbf{D}_{in})^{-1} \cdot (\mathbf{D}_{in}^T \otimes \mathbf{x}_{in}))} \quad (5)$$

[0047]

[0048] VBM 접근법을 이용할 경우, $\mathbf{x}_{in}$ 은 단일 벡터이고 $\mathbf{D}_a$ 는 2차원 어레이인 것이 이해될 것이다. SSM 접근법의 경우, $\mathbf{x}_{in}$ 은 시간 시퀀스 벡터의 어레이이고, $\mathbf{D}_a$ 는 시간 시퀀싱된 어레이의 집합이다. 이렇게 작성된 모델은 추정값을 생성하는데 이용된다. 예를 들어, 요청된 센서에 대한 추정치는 현재 시점에 대해(VBM 접근법이 이용될 경우) 또는 미래 시점에 대해(SSM 접근법이 이용될 경우) 얻어질 수 있다.

[0049] 모델 작성 모듈(124)은 또한 가지치기 알고리즘을 이용하여 임의의 불완전하게 수행하는 앙상블 모델을 제거할 수 있다. 일 양태에서의 가지치기 알고리즘은 전역 유사성(global similarity)이라 불리는 통계치를 이용하는데, 이것은 그 전체가 참조문헌으로 포함되는 미국 특허 6,859,739에 기술되어 있다.

[0050] 분포 생성 모듈(126)은 복수의 커널 회귀 모델을 이용하여, 요청된 센서 값의 적어도 하나의 분포를 생성한다. 이제 간단하게 도 2를 참조하여, 본 접근법에 의해 사용되는 통계 정보의 일례에 대해 설명한다. x축은 관심 센서에 대한 추정 값을 나타내는 다양한 점들을 갖는다. 각 점은 별개의 앙상블 모델로부터의 별개의 추정치이다. y축은 (x축 상에서) 소정의 간격에 있는 점들의 수를 나타낸다. 소정의 x축 간격에서의 빈도 또는 점의 개수의 플롯(202)이 얻어지고, 일 양태에서는 가우시안형(Gaussian-like) 분포인 것을 알 수 있다. 플롯(202)은 중앙값(206) 및 표준 편차(204)를 갖는다. 2개의 표준 편차가 예컨대 추정치 전체의 90 %를 나타낸다. 따라서, 중앙값 추정치는 일례에서 대략 3.8 +/- 1이다.

[0051] 분석 모듈(128)은 요청된 센서 값의 분포를 분석하여 적어도 하나의 분포의 중심의 척도 및 적어도 하나의 분포의 폭의 척도를 취득한다. 전술한 바와 같이, 분포 생성 모듈(126)은 추정치 점들을 얻는 모델 작성 모듈(124)에 의해 취득된 모델을 사용하여 추정치 점들의 분포를 산출한다. 일례로서, VBM 접근법을 이용하면, 다양한 모델이 추정 점들을 달성하는데 이용된다. 각각의 추정치 점은 사용자가 원하는 센서 값의 추정치를 나타낼 수 있다. 분석 모듈(128)은 몇몇 예를 들자면, 평균(즉, 추정치 전체의 합을 추정치의 수로 나눈 것), 중앙값, 및 표준 편차를 계산할 수 있다. 이 정보는 출력 인터페이스(116)를 통해 사용자에게 제공될 수 있다.

[0052] 이제 도 3을 참조하여, 추정치를 취득하기 위한 한 접근법을 설명한다. 단계 302에서, 엔티티 또는 프로세스와 연관된 물리적 파라미터를 나타내는 정보가 감지된다.

[0053] 단계 304에서, 감지된 정보가 현재 패턴에 또는 패턴의 현재 시퀀스에 수집된다. 단계 306에서, 현재 패턴 또는 패턴의 현재 시퀀스를 이력 데이터와 비교하여 베스트 매칭의 모집단을 얻는다.

[0054] 단계 308에서, 베스트 매칭의 모집단에 기초하여 복수의 커널 회귀 모델이 작성된다. 단계 310에서, 복수의 커널 회귀 모델을 이용하여 관심 센서에 대해 추정치의 적어도 하나의 분포가 생성된다. 단계 312에서, 관심 센서에 대한 추정치의 적어도 하나의 분포를 분석하여 적어도 하나의 추정치 분포의 중심의 척도 및 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 취득한다.

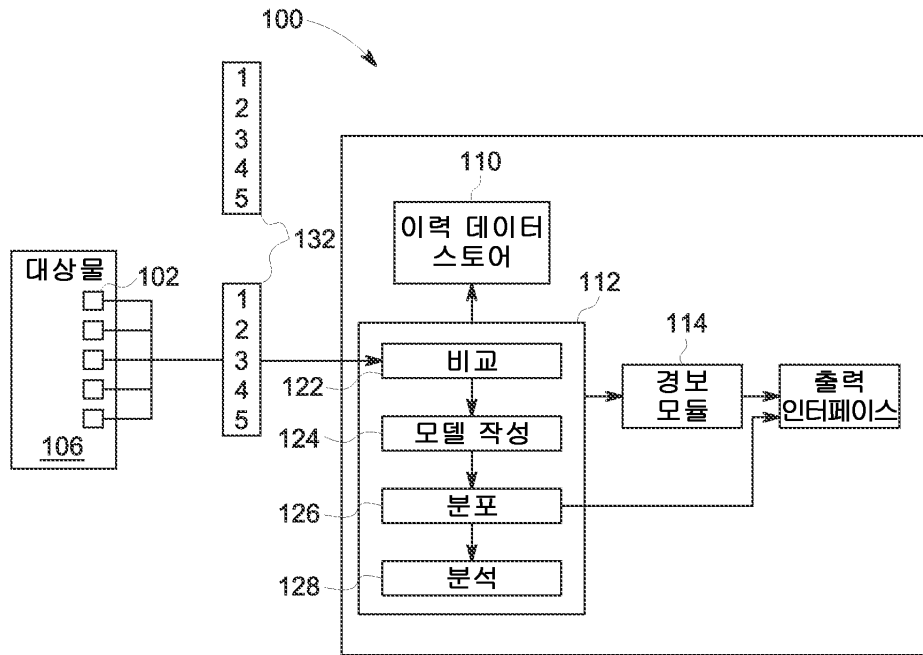
[0055] 이제 도 4를 참조하면, 추정치를 취득하기 위한 장치(400)는 인터페이스(402)와 프로세서(404)를 포함한다. 인

터페이스(402)는 입력부(406)와 출력부(408)를 포함하며, 입력부(406)는 엔티티 또는 프로세스와 연관된 물리적 파라미터를 나타내는 감지된 정보를 수신하도록 구성된다. 감지된 정보는 현재 패턴에 또는 패턴(410)의 현재 시퀀스에 수집된다.

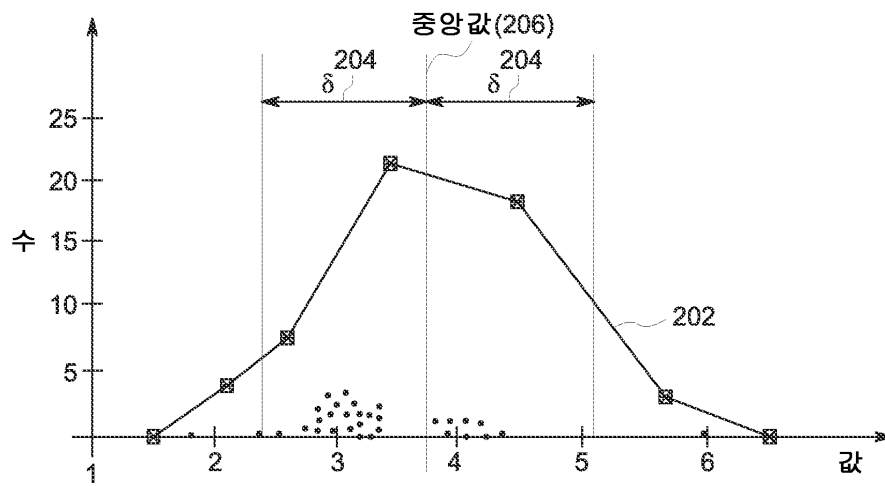
- [0056] 프로세서(404)는 인터페이스(402)에 연결된다. 프로세서(404)는 베스트 매칭의 모집단을 얻기 위해 현재 패턴 또는 패턴(410)의 현재 시퀀스를 메모리(414) 내의 이력 데이터(412)와 비교하도록 구성된다. 여기에서 사용하는 "베스트" 매칭(best match)은 소정의 기준, 표준, 기대값 또는 지침을 충족하거나 초과하는 매칭을 의미한다. 정확한 기준, 표준, 기대값 또는 지침은 특정 사용자 또는 시스템의 요구에 맞게 조정될 수 있다.
- [0057] 프로세서(404)는, 베스트 매칭의 모집단에 기초하여 복수의 커널 회귀 모델을 작성하고 복수의 커널 회귀 모델을 이용하여 관심 센서에 대한 추정치의 적어도 하나의 분포를 생성하도록 구성된다.
- [0058] 프로세서(404)는 또한, 관심 센서에 대한 추정치의 적어도 하나의 분포를 분석하여 적어도 하나의 추정치 분포의 중심의 척도 및 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 취득하도록 구성된다. 프로세서(404)는 적어도 하나의 추정치 분포의 중심의 척도 및 적어도 하나의 추정치 분포의 추정치 분포 폭의 척도를 출력부(408)에 제공한다.
- [0059] 기존의 접근법을 능가한 상업적 효과를 제공하는 본 접근법의 응용의 일례는 펌프를 이용한 석유 및 가스 추출에 관련된다. 유정 및 가스정 내에 그리고 전기 잠수정 펌프 상에 있는 다운홀 센서가 용기 온도, 용기 압력, 및 펌프 속도 등의 파라미터의 연속 측정을 제공하지만, 추출된 석유 및 가스의 양을 결정하는 데에 이용되는 주요 유정 성능 파라미터 중 어느 것도 제공하지 않는다. 체적 유량 및 워터-컷(water-cut)(즉, 유정으로부터 생산된 총 액체의 부피 대비 생산된 물의 비율) 등의 주요 성능 파라미터는 유정 테스트 시에 (잘해야) 불규칙한 간격으로 측정된다. 지속적인 센서 신호와 간헐적인 주요 성능 신호의 양상불 커널 회귀 모델을 작성함으로써, 체적 유량 및 워터-컷 파라미터는 (현재 시간에) 유정 테스트가 수행되지 않을 때 연관된 신뢰도 대역에 의해 추정될 수 있다.
- [0060] 본 접근법이 모델 트레이닝 백터의 선택을 랜덤화하는 것이 이해될 것이다. 즉, 주어진 모델 센서 세트에 대해, 센서 값을 포함하는 다양한 관찰 백터가 얻어진다. 그러나, 다른 예에서는, 사용되는 피처(예컨대, 사용되는 센서)가 랜덤화될 수도 있다. 즉, 특정 양상불 모델에서 변수로서 포함되는 센서가 랜덤하게 선택된다. 실제로, 하나의 양상불 모델은 제1 및 제2 센서로부터의 데이터를 이용할 수 있다. 제2 양상불 모델에서는, 다른 센서 그룹(제3 및 제4 센서)으로부터의 데이터가 이용될 수 있다. 제3 양상불 모델에서는, 제3 센서 그룹(예컨대, 제1 센서 및 제3 센서)으로부터의 데이터가 이용될 수 있다.
- [0061] 전술한 바와 같이, 본 접근법은 VBM 접근법을 이용하여 현재의 누락 추정치를 추론한다. 본 예에 있어서, 이것은 $\pm$ 범위를 가진 현재 값의 체적 유량일 수 있다. 다른 접근법에서는 SSM 모델에 따라 미래의 추정치가 얻어질 수 있다. 예컨대, 앞으로 2일과 3일 내의 체적 유량이 $\pm$ 범위로 추정될 수 있다.
- [0062] 다른 응용에 있어서, 본 접근법은 풍력 발전단지 및/또는 전체 풍력 발전단지의 개개의 터빈에 의해 제공되는 출력 전력의 예측치를 얻기 위해 풍력 발전단지에 구성되는 풍력 터빈에 적용될 수 있다. 이러한 측면에서, 발전단지 내의 다양한 터빈으로부터의 이력 풍력 데이터가 저장되어, 여기에 설명하는 모델을 작성하는데 이용될 수 있다. 본 접근법에 따르면, 주어진 날에, 풍속 또는 다른 센서 관독치가 풍력 발전단지 내의 소정의 터빈 또는 시점으로부터(예컨대, 오전 9시 및 오전 10시에 모든 센서로부터) 소정 회 취해질 수 있다. 현재의 접근법과 함께 SSM 모델링을 사용하여, 여러 모델이 작성되고, 이들 모델은 특정 터빈의 전력 출력 추정치를 생성하고/하거나 전체 풍력 발전단지의 전력 출력은 미래의 주어진 시간에 대해(예컨대, 풍력 발전단지에서 99 MW $\pm$ 9 MW의 전력을 생산할 같은 날 오전 11시), 또는 통계적 허용차와 함께 미래의 날에 대해(예컨대, 내일 오전 11시에 풍력 발전단지에서 101 MW $\pm$ 10 MW의 전력을 생산할 예정임) 얻어질 수 있다.
- [0063] 이들은 현재의 접근법이 채택되어 이용될 수 있는 두가지 응용례일 뿐임이 이해될 것이다. 다른 예도 가능하다.
- [0064] 본 발명의 바람직한 실시형태는 본 발명을 수행하기 위해 본 발명자들에게 알려진 최선의 양태를 포함하여 여기에 기술된다. 개시하는 실시형태는 예시적인 것일 뿐이며, 본 발명의 범위를 제한하는 것으로서 받아들여서는 안 되는 것이 이해되어야 한다.

도면

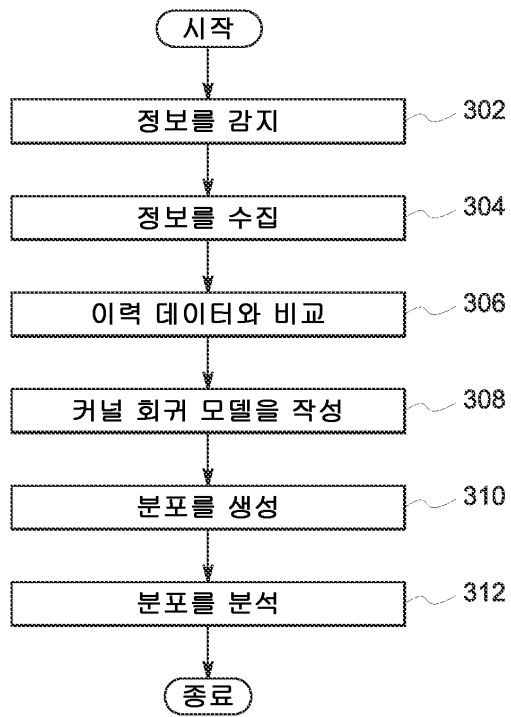
도면1



도면2



도면3



도면4

