



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0080111
(43) 공개일자 2018년07월11일

(51) 국제특허분류(Int. Cl.)
G06N 99/00 (2010.01) G06F 15/18 (2018.01)
G06N 3/08 (2006.01)
(52) CPC특허분류
G06N 99/005 (2013.01)
G06F 15/18 (2018.05)
(21) 출원번호 10-2017-0177880
(22) 출원일자 2017년12월22일
심사청구일자 없음
(30) 우선권주장
1020170000690 2017년01월03일 대한민국(KR)

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
(72) 발명자
오세원
대전광역시 서구 둔지로 75, 103동 705호 (둔산동, 파랑새아파트)
이연희
대전광역시 유성구 노은로 353, 304동 1701호(하기동, 송림마을아파트3단지)
(뒤틀면에 계속)
(74) 대리인
특허법인지명

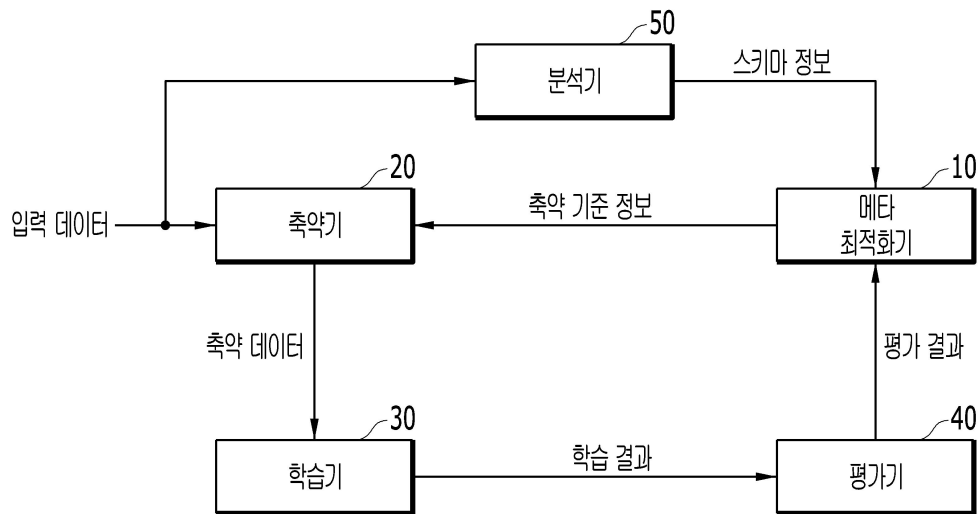
전체 청구항 수 : 총 17 항

(54) 발명의 명칭 자가 학습을 위한 데이터 메타 스케일링 장치 및 방법

(57) 요약

데이터 메타 스케일링 방법이 개시된다. 이 방법은, 기계 학습의 수행 과정에서 데이터를 표현할 수 있는 다양한 차원에서 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준을 최적화한다.

대표도 - 도1



- (52) CPC특허분류
G06N 3/08 (2013.01)
- (72) 발명자
배지훈
 세종특별자치시 새롬남로 18, 104동 1002호 (새롬동, 새뜸마을1단지)
- 강현중**
 경상남도 진주시 수곡면 원창길 116
- 권순현**
 인천광역시 서구 검암로 53, 105동 204호 (검암동, 풍림아이원2차)
- 김귀훈**
 대전광역시 서구 만년로 25, 105동 1002호 (만년동, 강변아파트)
- 김영민**
 대전광역시 유성구 전민로 30, 101호 (전민동, 아이캐슬)
- 김은주**
 대전광역시 유성구 대덕대로590번길 11-10, 701호 (도룡동, 더포엠)
- 김현재**
 인천광역시 부평구 굴포로 158, 502동 2001호 (삼산동, 행복한마을서해그랑블)

- 박홍규**
 대전광역시 서구 월평새뜸로4번길 53, 302호 (월평동, 채움빌라)
- 유재학**
 대전광역시 유성구 배울1로 35, 405동 1101호 (관평동, 대덕테크노밸리쌍용스윗닷홈4단지아파트)
- 이호성**
 대전광역시 유성구 배울1로 119, 1209동 1001호 (용산동, 대덕테크노밸리12단지아파트)
- 조성익**
 대전광역시 유성구 죽동로 321, 108동 501호 (죽동, 금성백조예미지)
- 김내수**
 대전광역시 대덕구 동춘당로 178, 104동 1303호 (법동, 보람아파트)
- 김선진**
 대전광역시 유성구 왕가봉로 23, 1101동 602호 (노은동, 열매마을아파트 11단지)
- 표철식**
 세종특별자치시 새롬남로 102, 1204동 1102호 (새롬동, 새뜸마을12단지)

이 발명을 지원한 국가연구개발사업

과제고유번호	CRC-15-05-ETRI
부처명	국가과학기술연구회
연구관리전문기관	국가과학기술연구회
연구사업명	융합연구사업
연구과제명	자가학습형 지식융합 슈퍼브레인 핵심기술개발
기 여 율	1/1
주관기관	한국전자통신연구원
연구기간	2015.12.01 ~ 2016.11.30

명세서

청구범위

청구항 1

프로세서가, 입력 데이터를 다른 속성에서 표현되도록 축약하는 규칙을 정의한 축약 기준 정보, 상기 축약 데이터에 대한 학습을 제한하는 규칙과 학습 성능의 평가 규칙을 정의한 학습 기준 정보 및 상기 축약 기준 정보를 최적화하기 위한 규칙을 정의한 지식 증강 기준 정보를 설정하는 단계;

상기 프로세서가, 상기 축약 기준 정보를 기반으로 상기 입력 데이터를 축약 데이터로 축약하는 단계;

상기 프로세서가, 상기 학습 기준 정보를 기반으로 상기 축약 데이터에 대해 학습을 수행하여 학습 모델을 생성하는 단계;

상기 프로세서가, 상기 학습 기준 정보를 기반으로 상기 학습 모델의 성능을 평가하여, 상기 축약 데이터의 적절성을 판단하는 단계; 및

상기 프로세서가, 상기 적절성을 판단한 결과에 따라서 상기 지식 증강 기준 정보를 기반으로 상기 축약 기준 정보를 갱신하는 지식 증강을 수행하는 단계;

를 포함하는 자가학습을 위한 데이터 메타 스케일링 방법.

청구항 2

제1항에서, 상기 설정하는 단계는,

다수의 속성으로 표현되는 상기 입력 데이터를 상기 다수의 속성 중에서 적어도 하나의 속성으로 표현되도록 축약하는 규칙을 정의한 상기 축약 기준 정보를 설정하는 단계를 포함하는 것인 자가학습을 위한 데이터 메타 스케일링 방법.

청구항 3

제1항에서, 상기 설정하는 단계는,

상기 입력 데이터가 다수의 속성으로 표현될 때, 상기 다수의 속성 중 어느 하나의 속성을 정의하는 데이터 차원을 나타내는 정보, 상기 입력 데이터의 샘플링 단위를 정의하는 윈도우를 나타내는 정보, 상기 윈도우의 종류를 나타내는 정보, 상기 윈도우의 크기를 나타내는 정보 및 상기 윈도우 내에서 대표값을 선택하는 기준을 나타내는 정보를 포함하도록 구성된 상기 축약 기준 정보를 설정하는 단계

를 포함하는 것인 자가학습을 위한 데이터 메타 스케일링 방법.

청구항 4

제1항에서, 상기 설정하는 단계는,

상기 입력 데이터의 종류를 나타내는 정보, 상기 학습 모델의 성능을 평가하기 위한 학습 신뢰도의 조건을 나타내는 정보, 상기 학습 신뢰도의 계산 방법을 나타내는 정보 및 상기 축약 데이터에 대한 학습의 반복 횟수를 제한하는 상기 학습의 조기 종료 조건을 나타내는 정보를 포함하도록 구성된 상기 학습 기준 정보를 설정하는 단계

를 포함하는 것인 자가학습을 위한 데이터 메타 스케일링 방법.

청구항 5

제1항에서, 상기 설정하는 단계는,

상기 축약 기준 정보의 변동 횟수를 나타내는 정보, 상기 축약 기준 정보의 변동 요소를 나타내는 정보, 상기 변동 요소의 변동 범위를 나타내는 정보, 상기 축약 데이터에 대해 학습을 수행하는 과정에서 발생한 학습 이력의 누적 횟수를 나타내는 정보를 포함하도록 구성된 상기 지식 증강 기준 정보를 설정하는 단계

를 포함하는 것인 자가학습을 위한 데이터 메타 스케일링 방법.

청구항 6

제5항에서, 상기 변동 요소는,

상기 입력 데이터의 샘플링 단위를 정의하는 윈도우와 관련된 정보인 것인 데이터 메타 스케일링 방법.

청구항 7

제6항에서, 상기 윈도우와 관련된 정보는,

상기 윈도우의 크기 및 윈도우들 사이의 간격을 나타내는 정보들을 포함하는 것인 데이터 메타 스케일링 방법.

청구항 8

제1항에서, 상기 축약하는 단계는,

상기 입력 데이터가 다수의 속성으로 표현되고, 상기 다수의 속성을 다수의 데이터 차원으로 각각 정의 할 때, 각 데이터 차원에서 상기 입력 데이터를 상기 입력 데이터의 대표값으로 샘플링하는 제1 프로세스, 상기 입력 데이터를 상기 다수의 데이터 차원에서 선택된 적어도 하나의 데이터 차원으로 변경하는 제2 프로세스 및 상기 제1 프로세스와 상기 제2 프로세스가 결합된 제3 프로세스 중에서 어느 하나의 프로세스에 의해 상기 입력 데이터를 축약 데이터로 축약하는 단계인 것인 데이터 메타 스케일링 방법.

청구항 9

제8항에서, 상기 제1 프로세스는,

상기 입력 데이터를 상기 입력 데이터의 대표값으로 주기적으로 샘플링하는 프로세스;

상기 입력 데이터를 상기 입력 데이터의 대표값으로 비주기적으로 샘플링하는 프로세스;

상기 입력 데이터의 샘플링 단위를 정의하는 윈도우가 복수이고, 복수의 윈도우가 중복되지 않은 상태에서 각 윈도우에서 대표값을 선택하는 고정 윈도우 기반의 샘플링 프로세스; 및

상기 복수의 윈도우가 중복된 상태에서 각 윈도우에서 대표값을 선택하는 무빙 윈도우 기반의 샘플링 프로세스를 포함하는 것인 데이터 메타 스케일링 방법.

청구항 10

제1항에서, 상기 지식 증강을 수행하는 단계는,

상기 학습 모델의 성능을 평가하기 위해 계산된 학습 신뢰도가 학습 기준 정보에서 정의하는 상기 학습 성능의 평가 규칙에서 규정한 조건을 만족하지 못하는 경우, 상기 지식 증강 기준 정보에서 정의하는 상기 축약 기준 정보의 변동 요소를 나타내는 정보 및 상기 변동 요소의 변동 범위에 따라 상기 축약 기준 정보를 변경하는 단

계; 및

상기 변경된 축약 기준 정보에 따라 축약된 상기 축약 데이터에 대해 학습을 수행하여 생성된 학습 모델의 성능이 상기 학습 기준 정보에서 규정하는 조건을 만족하면, 상기 변경된 축약 기준 정보를 최적의 축약 기준 정보로 갱신하는 단계

를 포함하는 것인 데이터 메타 스케일링 방법.

청구항 11

입력 데이터를 다른 속성에서 표현되도록 축약하는 규칙을 정의한 축약 기준 정보, 상기 축약 데이터에 대한 학습을 제한하는 규칙과 학습 성능의 평가 규칙을 정의한 학습 기준 정보 및 상기 축약 기준 정보를 최적화하기 위한 규칙을 정의한 지식 증강 기준 정보를 설정하는 메타 최적화기;

상기 축약 기준 정보를 기반으로 상기 입력 데이터를 축약 데이터로 축약하는 축약기;

상기 학습 기준 정보를 기반으로 상기 축약 데이터에 대해 학습을 수행하여 학습 모델을 생성하는 학습기; 및

상기 학습 기준 정보를 기반으로 상기 학습 모델의 성능을 평가하여, 상기 축약 데이터의 적절성을 판단하는 평가기;를 포함하고,

상기 메타 최적화기는,

상기 적절성을 판단한 결과에 따라서 상기 지식 증강 기준 정보를 기반으로 상기 축약 기준 정보를 갱신하는 지식 증강을 수행하는 것인 자가학습을 위한 데이터 메타 스케일링 장치.

청구항 12

제11항에서, 상기 메타 최적화기는,

다수의 속성으로 표현되는 상기 입력 데이터를 상기 다수의 속성 중에서 적어도 하나의 속성으로 표현되도록 축약하는 규칙을 정의한 상기 축약 기준 정보를 설정하는 것인 자가학습을 위한 데이터 메타 스케일링 장치.

청구항 13

제11항에서, 상기 메타 최적화기는,

상기 입력 데이터가 다수의 속성으로 표현될 때, 상기 다수의 속성 중 어느 하나의 속성을 정의하는 데이터 차원을 나타내는 정보, 상기 입력 데이터의 샘플링 단위를 정의하는 윈도우를 나타내는 정보, 상기 윈도우의 종류를 나타내는 정보, 상기 윈도우의 크기를 나타내는 정보 및 상기 윈도우 내에서 대표값을 선택하는 기준을 나타내는 정보를 포함하도록 구성된 상기 축약 기준 정보를 설정하는 것인 자가학습을 위한 데이터 메타 스케일링 장치.

청구항 14

제11항에서, 상기 메타 최적화기는,

상기 입력 데이터의 종류를 나타내는 정보, 상기 학습 모델의 성능을 평가하기 위한 학습 신뢰도의 조건을 나타내는 정보, 상기 학습 신뢰도의 계산 방법을 나타내는 정보 및 상기 축약 데이터에 대한 학습의 반복 횟수를 제한하는 상기 학습의 조기 종료 조건을 나타내는 정보를 포함하도록 구성된 상기 학습 기준 정보를 설정하는 것인 자가학습을 위한 데이터 메타 스케일링 장치.

청구항 15

제11항에서, 상기 메타 최적화기는,

상기 축약 기준 정보의 변동 횟수를 나타내는 정보, 상기 축약 기준 정보의 변동 요소를 나타내는 정보, 상기 변동 요소의 변동 범위를 나타내는 정보, 상기 축약 데이터에 대해 학습을 수행하는 과정에서 발생하는 학습 이력의 누적 횟수를 나타내는 정보를 포함하도록 구성된 상기 지식 증강 기준 정보를 설정하는 것인 자가학습을 위한 데이터 메타 스케일링 장치.

청구항 16

제15항에서, 상기 변동 요소는,

상기 입력 데이터의 샘플링 단위를 정의하는 윈도우와 관련된 정보인 것인 데이터 메타 스케일링 장치.

청구항 17

제11항에서, 상기 메타 최적화기는,

상기 학습 모델의 성능이 상기 학습 성능의 평가 규칙에서 규정한 조건을 만족하지 못하는 경우, 상기 지식 증강 기준 정보에서 정의하는 상기 축약 기준 정보의 변동 요소 및 상기 변동 요소의 변동 범위에 따라 상기 축약 기준 정보를 변경하고, 상기 변경된 축약 기준 정보에 따라 축약된 상기 축약 데이터에 대해 학습을 수행하여 생성된 상기 학습 모델의 성능이 상기 학습 기준 정보에서 규정하는 조건을 만족하면, 상기 변경된 축약 기준 정보를 상기 갱신된 축약 기준 정보로서 저장소에 저장하여 지식 증강을 수행하는 것인 데이터 메타 스케일링 장치.

발명의 설명

기술 분야

[0001] 본 발명은 자가 학습을 위한 데이터 메타 스케일링 장치 및 방법에 관한 것으로서, 더욱 상세하게는 기계 학습 모델의 학습에 이용되는 입력 데이터의 처리 기술에 관한 것이다.

배경 기술

[0003] 기계 학습(Machine Learning, ML)은 수집된 데이터를 분류하거나 수집된 데이터의 특성을 나타내는 모델을 학습하는 용도로 널리 활용되고 있다. 기계 학습과 관련해, 다양한 기술이 개발되고 있으며, 이러한 기계 학습에서 최적의 분류 성능 또는 학습 성능을 도출하기 위해서는, 수집된 데이터를 그대로 이용하는 것보다 기계 학습 알고리즘 또는 도출하고자 하는 목표에 따라서 적절하게 축약하여 학습하는 것이 바람직하다. 즉, 여러 사물을 통해 대용량 데이터를 지속적으로 수집하는 환경에서는 데이터의 활용 목적이나 주변 환경에 맞추어 적절하게 축약된 데이터를 학습하도록 기계 학습 시스템을 제어하는 것이 매우 중요하다. 그러나 아직까지 적절하게 축약된 데이터를 기반으로 학습 과정을 수행하는 기계 학습 시스템에 대한 개발이 미진한 상태이다.

발명의 내용

해결하려는 과제

[0005] 본 발명에서 해결하고자 하는 과제는 기계 학습의 수행 과정에서 데이터를 표현할 수 있는 다양한 차원에서 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있는 자가 학습을 위한 데이터 메타 스케일링 장치 및 방법을 제공하는 데 있다.

과제의 해결 수단

[0007] 상술한 과제 해결 수단으로서, 본 발명의 자가학습을 위한 데이터 메타 스케일링 방법은 프로세서가, 입력 데이터를 다른 속성에서 표현되도록 축약하는 규칙을 정의한 축약 기준 정보, 상기 축약 데이터에 대한 학습을 제한하는 규칙과 학습 성능의 평가 규칙을 정의한 학습 기준 정보 및 상기 축약 기준 정보를 최적화하기 위한 규칙을 정의한 지식 증강 기준 정보를 설정하는 단계; 상기 프로세서가, 상기 축약 기준 정보를 기반으로 상기 입력 데이터를 축약 데이터로 축약하는 단계; 상기 프로세서가, 상기 학습 기준 정보를 기반으로 상기 축약 데이터에 대해 학습을 수행하여 학습 모델을 생성하는 단계; 상기 프로세서가, 상기 학습 기준 정보를 기반으로 상기 학습 모델의 성능을 평가하여, 상기 축약 데이터의 적절성을 판단하는 단계; 및 상기 프로세서가, 상기 적절성을 판단한 결과에 따라서 상기 지식 증강 기준 정보를 기반으로 상기 축약 기준 정보를 갱신하는 지식 증강을 수행하는 단계;를 포함한다.

[0008] 본 발명의 자가학습을 위한 데이터 메타 스케일링 장치는 입력 데이터를 다른 속성에서 표현되도록 축약하는 규칙을 정의한 축약 기준 정보, 상기 축약 데이터에 대한 학습을 제한하는 규칙과 학습 성능의 평가 규칙을 정의한 학습 기준 정보 및 상기 축약 기준 정보를 최적화하기 위한 규칙을 정의한 지식 증강 기준 정보를 설정하는 메타 최적화기; 상기 축약 기준 정보를 기반으로 상기 입력 데이터를 축약 데이터로 축약하는 축약기; 상기 학습 기준 정보를 기반으로 상기 축약 데이터에 대해 학습을 수행하여 학습 모델을 생성하는 학습기; 및 상기 학습 기준 정보를 기반으로 상기 학습 모델의 성능을 평가하여, 상기 축약 데이터의 적절성을 판단하는 평가기;를 포함하고, 상기 메타 최적화기는, 상기 적절성을 판단한 결과에 따라서 상기 지식 증강 기준 정보를 기반으로 상기 축약 기준 정보를 갱신하는 지식 증강을 수행한다.

발명의 효과

[0010] 본 발명에 따르면, 기계 학습에서 최적 성능을 달성할 수 있도록, 기계 학습을 수행하고자 하는 데이터에 대하여 다양한 차원에서 데이터 축약 프로세스를 수행함으로써, 자가 학습을 구성하는 절차를 자동화할 수 있으며, 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있다.

[0011] 더하여, 본 발명은 스키마 정보(schema information)를 참조하여 축약 기준 정보를 갱신하기 위한 기준 및 조건을 정의한 지식 증강 기준 정보를 설정하고, 이를 토대로 서로 다른 복수의 축약 기준 정보를 설정하여 데이터를 축약하고, 축약된 데이터를 서로 다른 복수의 기계 학습에 병렬로 적용하여 평가함으로써 다양한 축약 기준 정보를 고려한 학습 이력의 생성과 저장이 가능하다.

[0012] 더하여, 본 발명은 입력 데이터 정보, 스키마 정보, 학습 모델 정보, 축약 기준 정보, 축약 데이터 정보, 학습 기준 정보, 학습 데이터 정보, 학습 모델 정보, 학습 결과 정보, 지식 증강 기준 정보 등을 포함하는 학습 이력 정보를 누적하여 저장하고, 저장된 학습 이력 정보를 바탕으로 최적의 축약 기준 정보를 자동 설정하는 지식 증강을 통해 축약 기준 정보를 최적화할 수 있다.

[0013] 더하여, 본 발명에 따른 데이터 메타 스케일링 기술은, 사물인터넷(IoT) 및 만물인터넷(IoE) 환경 등에서 수집되는 다양한 종류의 데이터를 표현할 수 있는 다양한 차원의 축약을 수행함으로써, 원본 데이터를 다른 구조로 변환할 수 있을 뿐만 아니라, 축약된 정보를 기반으로 원본 데이터에 새로운 속성을 추가하여 확장할 수 있다.

도면의 간단한 설명

[0015] 도 1은 본 발명의 제1 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 장치를 도시한 블록도이다.
 도 2는 본 발명의 제1 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 방법을 도시한 순서도이다.
 도 3a 내지 3c는 본 발명의 실시 예에 따른 데이터 축약에서 단일 차원 기반의 샘플링을 설명하기 위한 도면들이다.
 도 4는 본 발명의 실시 예에 따른 데이터 축약에서 복합 차원 기반의 샘플링을 설명하기 위한 도면이다.
 도 5는 본 발명의 다른 실시 예에 따른 데이터 축약에서 복합 차원 기반의 샘플링을 설명하기 위한 도면
 도 6a 내지 6c는 본 발명의 일 실시 예에 따른 스키마 정보에 포함된 축약 기준 정보, 학습 기준 정보 및 지식 증강 기준 정보의 데이터 구조를 나타낸 도면들이다.

도 7은 본 발명에 일 실시 예에 따른 스키마 정보를 온톨로지로 표현한 도면이다.

도 8은 본 발명의 제2 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 장치를 도시한 블록도이다.

도 9는 본 발명의 제3 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 장치를 도시한 블록도이다.

도 10은 도 1에 도시한 데이터 메타 스케일링 장치를 교통 정보 예측 시나리오에 적용하는 예를 설명하기 위한 도면이다.

도 11a 내지 11c는 본 발명의 일 실시 예에 따른 최적의 축약 기준을 찾는 지식 증강 프로세스를 도식적으로 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0016] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세하게 설명한다. 사용되는 용어들은 실시예에서의 기능을 고려하여 선택된 용어들로서, 그 용어의 의미는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 후술하는 실시예들에서 사용된 용어의 의미는, 본 명세서에 구체적으로 정의된 경우에는 그 정의에 따르며, 구체적인 정의가 없는 경우는 당업자들이 일반적으로 인식하는 의미로 해석되어야 할 것이다.
- [0017] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.
- [0018] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0019] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0020] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0021] 본 발명에 따른 자가 학습을 위한 데이터 메타 스케일링 장치 및 방법의 구성 및 기능은, 하나 이상의 컴퓨터 판독 가능한 명령어를 포함하도록 구성된 프로그램 모듈로 구현될 수 있다.
- [0022] 프로그램 모듈은 메모리 등의 기록매체에 저장된 후, 프로세서에 의해 로딩되어 실행됨으로써 본 발명에서 설명하는 특정 기능을 수행할 수 있다. 컴퓨터 판독가능 명령어는, 예를 들면, 범용 컴퓨터 시스템 또는 특수 목적 컴퓨터 시스템이 특정 기능 또는 기능의 그룹을 수행하도록 하는 명령어 및 데이터를 포함한다.
- [0023] 컴퓨터 실행가능 명령어는, 예를 들면, 어셈블리어, 또는 소스코드와 같은 이진, 중간 포맷 명령어일 수 있다. 즉, 본 발명에 따른 자가 학습을 위한 데이터 메타 스케일링 장치 및 방법은, 컴퓨터 프로그램을 포함하는 소프트웨어 또는 컴퓨터 시스템과 같이 메모리 및 프로세서 등을 포함하는 하드웨어 및 상기 하드웨어에 탑재되어 상기 하드웨어에 의해 실행되는 소프트웨어의 조합으로 구현될 수 있다.
- [0024] 본 발명에 따른 방법을 실행하는 컴퓨터 프로그램은 컴파일되거나 해석된 언어나 선형적(transcendental) 혹은 절차적 언어(procedural language)를 포함하는 프로그래밍 언어의 어떠한 형태로도 작성될 수 있으며, 독립형 프로그램이나 모듈, 컴포넌트, 서브루틴 혹은 컴퓨터 환경에서 사용하기에 적합한 다른 유닛을 포함하여 어떠한 형태로도 전개될 수 있다.

- [0025] 컴퓨터 프로그램은 파일 시스템의 파일에 반드시 대응하는 것은 아니다. 프로그램은 요청된 프로그램에 제공되는 단일 파일 내에, 혹은 다중의 상호 작용하는 파일(예컨대, 하나 이상의 모듈, 하위 프로그램 혹은 코드의 일부를 저장하는 파일) 내에, 혹은 다른 프로그램이나 데이터를 보유하는 파일의 일부(예컨대, 마크업 언어 문서 내에 저장되는 하나 이상의 스크립트) 내에 저장될 수 있다.
- [0026] 더하여, 상기 컴퓨터 프로그램은 하나의 사이트에 위치하거나 복수의 사이트에 걸쳐서 분산되어 통신 네트워크에 의해 상호 접속된 다중 컴퓨터나 하나의 컴퓨터 상에서 실행되도록 전개될 수 있다.
- [0027] 이러한 컴퓨터 프로그램을 저장하기에 적합한 컴퓨터로 판독 가능한 매체는, 예컨대 EPROM, EEPROM 및 플래시 메모리 장치와 같은 반도체 메모리 장치, 예컨대 내부 하드디스크나 외장형 디스크와 같은 자기 디스크, 자기광학 디스크 및 CD-ROM과 DVD-ROM 디스크를 포함하여 모든 형태의 비휘발성 메모리, 매체 및 메모리 장치를 포함한다. 프로세서와 메모리는 특수 목적의 논리 회로에 의해 보충되거나, 그것에 통합될 수 있다.
- [0028] 아울러, 본 발명에 따른 데이터 메타 스케일링 장치 및 방법은, 기계 학습 시스템에 적용되어, 기계 학습을 수행하는 과정에서, 다수의 속성으로 표현될 수 있는 입력 데이터에 대해서 스키마 정보를 기반으로 축약 기준 정보를 설정한다.
- [0029] 이에 따라서 본 발명에 따른 데이터 메타 스케일링 장치 및 방법은, 축약된 데이터에 대해 학습을 수행하고, 학습 결과를 이용하여 평가함으로써 기계 학습에 대해 최적의 성능을 도출할 수 있는 축약 데이터를 제공할 수 있다.
- [0030] 이러한 본 발명의 구성 및 동작을 다양한 실시 예를 기반으로 설명한다.
- [0031] 도 1은 본 발명의 제1 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 장치를 도시한 블록도이다.
- [0032] 본 발명의 제1 실시 예에 따른 데이터 메타 스케일링 장치는, 데이터 입력, 스키마 정보 추출, 데이터 축약, 모델 학습, 학습 이력 저장, 학습 이력 분석, 지식 증강의 절차를 자동화하는 프로세스를 수행할 수 있다. 자가 학습은 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있는 반복 가능한 학습 과정으로 정의한다.
- [0033] 본 발명의 제1 실시 예에 따른 데이터 메타 스케일링 장치는, 입력된 데이터 또는 사용자 입력에서 스키마 정보를 추출하고, 추출된 스키마 정보를 기반으로 축약 기준 정보, 학습 기준 정보 및 지식 증강 기준 정보를 구축하여 자가 학습 수행을 위한 준비를 완료한다.
- [0034] 다음으로, 본 발명의 제1 실시 예에 따른 데이터 메타 스케일링 장치는, 축약 기준 정보에서 규정하는 축약 기준 또는 축약 규칙을 기반으로 데이터 축약을 수행하고, 학습 기준 정보에서 규정하는 학습 기준을 기반으로 축약된 데이터를 적절히 표현할 수 있는 모델에 대한 학습을 수행한다. 학습 수행은 지식 증강 기준을 기반으로 반복 수행될 수 있으며, 학습 결과는 학습 이력으로 자동 저장된다.
- [0035] 지식 증강 기준 정보에서 규정하는 지식 증강 기준을 만족하도록 학습 이력이 충분히 저장되면, 본 발명의 제1 실시 예에 따른 데이터 메타 스케일링 장치는 학습 이력을 분석하여 축약 기준에 대한 최적화를 수행한다.
- [0036] 이러한 과정을 통해 자가 학습을 구성하는 절차를 자동화할 수 있으며, 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있다.
- [0037] 도 1을 참조하면, 본 발명의 제1 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 장치는, 메타 최적화기(10)와, 축약기(20)와, 학습기(30), 평가기(40) 및 분석기(50)를 포함한다.
- [0038] 메타 최적화기(10)는 입력 데이터의 스키마(schema) 정보를 참조하여 축약 기준 정보, 학습 기준 정보, 및 지식 증강 기준 정보를 설정하는 프로세스를 수행한다. 스키마 정보는 입력 데이터의 메타 데이터를 분석하여 획득할 수 있다. 메타 데이터는 입력 데이터의 특정 영역에 포함될 수 있다. 메타 데이터는 입력 데이터의 속성을 설명하는 데이터일 수 있다.
- [0039] 스키마 정보는 사용자 입력에 의해 제공될 수 있다. 입력 데이터는, 다수의 속성 정보로 이루어진 것으로, 연속적인 스트림 형태 또는 아카이브(archive) 형태로 제공될 수 있다. 예를 들어, 입력 데이터는 사물 인터넷 서비스 환경에서 센싱 기기 등과 같은 다양한 사물 기기로부터 수집되는 데이터들일 수 있다.
- [0040] 축약기(20)는 메타 최적화기(10)에서 설정한 축약 기준 정보를 이용하여 입력 데이터를 축약하는 프로세스를 수행한다. 입력 데이터는 다양한 사물 기기로부터 직접 입력되거나 데이터 저장소로부터 입력될 수 있다. 데이터

의 입력은 실제 데이터의 물리적인 입력과 데이터가 존재하는 논리적 위치 정보의 입력을 포함한다. 여기서, 논리적 위치 정보는, 예를 들면, URL 정보일 수 있다.

- [0041] 학습기(30)는 메타 최적화기(10)에서 설정한 학습 기준 정보를 이용하여 축약기(200)에서 축약한 축약 데이터에 대한 기계 학습을 수행한다. 기계 학습의 종류 또는 기계 학습 수행에 필요한 초매개변수(hyperparameter)의 특성 등을 한정하는 것은 본 발명의 요지를 벗어나는 것이므로, 이를 한정하지 않는다. 즉, 본 발명은 기계 학습 수행에 필요한 초매개변수의 특성에 관계없이 모든 종류의 기계 학습에 적용될 수 있으며, 이는 당업자라면 이하의 설명으로부터 충분히 이해할 수 있을 것이다. 한편, 학습기(30)는 축약 데이터 및 입력 데이터를 모두 이용하여 기계 학습을 수행할 수도 있다. 이는 데이터 축약을 통해 추출된 새로운 속성을 입력 데이터에 추가하여 입력 데이터를 확장하고, 확장된 입력 데이터로 학습을 수행할 수 있음을 의미한다.
- [0042] 평가기(40)는 메타 최적화기(10)에서 설정한 학습 기준 정보를 토대로 학습 과정이나 학습 결과가 학습 기준을 만족하는지 여부를 판단하고, 그 판단 결과를 기반으로 데이터 축약의 적절성을 평가하는 프로세스를 수행한다.
- [0043] 분석기(50)는 입력 데이터에 포함된 메타 데이터 또는 입력 데이터와 함께 제공되는 메타 데이터를 분석하여 입력 데이터의 스키마 정보를 추출한다.
- [0044] 메타 최적화기(10)는 평가기(40)의 평가 결과 정보를 바탕으로 축약 기준 정보의 변경 또는 지식 증강을 수행한다.
- [0045] 학습 과정이나 학습 결과가 학습 기준 정보에서 규정한 학습 기준을 만족하지 못하는 경우, 메타 최적화기(10)는 지식 증강 기준을 바탕으로 축약 기준 정보를 변경하는 프로세스를 수행한다. 반대로 학습 과정이나 학습 결과가 학습 기준을 만족하는 경우, 메타 최적화기(10)는 학습 결과를 학습 이력으로 저장부(50)에 자동 저장하는 프로세스를 통해 지식 증강 프로세스를 시작한다.
- [0046] 메타 최적화기(10)는 지식 증강 기준 정보에서 규정한 지식 증강 기준을 만족하도록 학습 이력이 충분히 저장되면, 저장된 학습 이력을 분석하여 축약 기준에 대한 최적화를 수행하는 프로세스를 수행한다. 이러한 프로세스를 통해 자가 학습을 구성하는 절차를 자동화할 수 있으며, 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있다.
- [0048] 도 2는 본 발명의 제1 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 방법을 도시한 순서도이다.
- [0049] 도 2를 참조하면, 먼저, 단계 S100에서, 입력 데이터가 사물 기기 또는 데이터 저장소로부터 메타 최적화기(10)로 입력되는 프로세스가 수행된다.
- [0050] 다음으로, 단계 S200에서, 메타 최적화기(10)가 입력 데이터에 포함된 메타 데이터를 분석하여(또는 파싱하여), 상기 입력 데이터의 스키마 정보를 추출하고, 추출된 스키마 정보를 기반으로 축약 기준 정보, 학습 기준 정보, 지식 증강 기준 정보를 구축하는 프로세스가 수행된다.
- [0051] 다음으로, 단계 S300에서, 축약기(20)에서 축약 기준 정보를 이용하여 입력 데이터를 축약하는 프로세스가 수행된다. 축약된 데이터는 학습기(30)로 실시간 스트림 또는 배치(batch) 방식으로 직접 제공될 수 있다. 다르게, 축약된 데이터가 저장매체에 저장되고, 축약기(20)가 그 저장 주소를 학습기(30)에 통지하는 방식으로 축약된 데이터의 제공을 대신할 수 있다. 이 경우, 학습기(30)는 저장 주소로 저장매체에 액세스하여 축약 기준 정보를 읽어들이 수 있다.
- [0052] 다음으로, 단계 S400에서, 학습기(30)에서 축약된 데이터를 적절히 표현할 수 있는 모델에 대한 학습을 수행하여 학습 모델을 생성하는 프로세스가 수행된다. 이 때, 학습기(30)는 학습 기준 정보를 바탕으로 학습을 수행한다.
- [0053] 다음으로, 단계 S500에서, 평가기(40)에서 학습 결과가 학습 기준 정보에서 정의하는 학습 기준을 만족하는지 여부를 판단하는 프로세스가 수행된다.
- [0054] 학습 결과가 학습 기준을 만족하지 못하는 경우, 단계 S600에서, 메타 최적화기(10)가 지식 증강 기준 정보에서 정의하는 지식 증강 기준을 기반으로 축약 기준 정보를 갱신하는 프로세스가 수행된다.
- [0055] 반대로, 학습 결과가 학습 기준을 만족하는 경우, 단계 S700에서, 메타 최적화기(10)가 지식 증강 기준을 만족하도록 학습 이력이 충분히 저장된 후, 충분히 저장된 학습 이력을 분석하여 축약 기준에 대한 최적화를 수행한다.

는 프로세스가 수행된다. 이러한 지식 증강 프로세스를 통해, 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있다.

[0056] 본 발명에서 입력 데이터는 다양한 속성을 갖는다. 이러한 다양한 속성을 표현하기 위해, 본 발명에서는 '데이터 차원'이라는 용어가 정의된다. 데이터 차원은 데이터를 표현할 수 있는 속성으로 정의한다.

[0058] **데이터 차원의 예**

[0059] 특정 시간 간격 또는 불특정 시간 간격으로 수집되는 데이터는 시간적 속성으로 표현할 수 있다. 따라서, 시간적 속성으로 표현할 수 있는 데이터의 차원은 '시간'이 된다.

[0060] 경/위도 좌표(Latitude and longitude coordinates), 주소정보(address information), 우편번호, IP의 서버넷 등과 같은 데이터는 물리적 또는 논리적 위치를 나타내는 공간적 속성으로 표현할 수 있다. 따라서, 공간적 속성으로 표현할 수 있는 데이터의 차원은 '공간'이 된다.

[0061] 색깔을 나타내는 데이터는 색도(hue), 채도(saturation), 및 명도(intensity)와 같은 속성으로 표현할 수 있다. 따라서, 색을 표현하는 데이터의 차원은 색도(hue), 채도(saturation) 또는 명도(intensity)가 된다.

[0062] 물질을 나타내는 데이터는 경도(hardness), 밀도(density), 비중(specific gravity), 전기전도도(conductivity) 등과 같은 물질의 고유 속성으로 표현할 수 있다. 따라서, 물질을 표현하는 데이터의 차원은 경도(hardness), 밀도(density), 비중(specific gravity), 전기 전도도(conductivity) 등일 수 있다.

[0063] 주파수에 따라 변환하는 데이터는 주파수를 데이터 차원으로 정의할 수 있다.

[0064] 거주지, 직장, 상가, 1층 등 사회적으로 부여되는 의미 범주에 따라 정의되는 데이터는 의미 범주를 데이터 차원으로 정의할 수 있다.

[0065] 사용자 집단이 어떤 서비스에 대해 평가한 결과를 나타내는 데이터의 차원은 선호도, 유효성 등이 될 수 있다.

[0066] 이동 카메라에서 촬영한 동영상은 촬영 위치, 촬영 시간 등을 데이터 차원으로 정의할 수 있다. 이때, 촬영 위치는 3차원 공간에서 XYZ 좌표로 표현할 수 있으므로, 3개의 데이터 차원으로 세분화될 수 있다.

[0067] 이와 같이 모든 데이터는 각자의 속성에 의해 다양한 차원으로 표현될 수 있으므로, 본 발명에서는 그 차원을 결정하는 기준에 대해서는 한정하지 않는다.

[0069] **데이터의 축약**

[0070] 본 발명의 실시 예에 따른 데이터 축약은, 임의의 데이터가 임의의 데이터 차원으로 표현될 때, 상기 임의의 데이터 차원에서 상기 임의의 데이터를 샘플링 하는 프로세스로 정의할 수 있다.

[0071] 또한, 본 발명의 실시 예에 따른 데이터 축약은 임의의 데이터가 갖는 데이터 차원을 다른 데이터 차원으로 변경하는 프로세스로 정의할 수 있다. 이러한 차원 변경은 데이터를 표현할 수 있는 범위의 축소를 의미한다. 경우에 따라, 차원 변경은 데이터를 표현할 수 있는 범위의 확대를 의미할 수도 있다.

[0072] 이와 같이, 본 발명의 실시 예에 따른 데이터 축약은 다양한 차원에서의 샘플링, 차원 변환 및 상기 샘플링과 상기 차원 변환을 결합하는 프로세스 중에서 어느 하나일 수 있으며, 이러한 프로세스를 통해 데이터 개수를 축소하는 과정으로 정의할 수 있다.

[0074] **데이터의 축약에 따른 샘플링**

[0075] 샘플링은 미리 정해진 기준에 따라 한 개 이상의 데이터 차원에서 대표값을 선택하는 프로세스일 수 있다.

[0076] 이러한 샘플링은 단일 차원 기반의 샘플링과 복합 차원 기반의 샘플링을 포함할 수 있다. 단일 차원 기반의 샘플링은 단일 데이터 차원에서 대표값을 선택하는 프로세스일 수 있다. 복합 차원 기반의 샘플링은 두 개 이상의 데이터 차원에서 대표값을 각각 선택하는 프로세스일 수 있다.

[0077] **A. 단일 차원 기반의 샘플링**

[0078] 단일 차원 기반의 샘플링 프로세스는 주기적 샘플링 프로세스, 비주기적 샘플링 프로세스, 고정 윈도우(fixed window) 기반의 샘플링 프로세스 및 무빙 윈도우(moving window) 기반의 샘플링 프로세스를 포함할 수 있다.

[0079] 주기적 샘플링 프로세스는 데이터 차원에서 주어진 윈도우에서 대표값을 주기적으로 선택하는 것으로, 예를 들면, 시간 차원에서 표현되는 데이터에 대해 5분 간격으로 주어진 윈도우에서 대표값을 특정 기준으로 선택하는 과정일 수 있다. 여기서, 윈도우는 샘플링 단위로 해석될 수 있다.

[0080] 비주기적 샘플링 프로세스는 주어진 윈도우에서 대표값을 비주기적으로 선택하는 것으로, 예를 들면, 데이터의 값이 미리 정해진 값 이상인 경우에 대해 주어진 윈도우에서 대표값을 특정 기준으로 선택하는 과정이거나 임의의 공간에서 온도 센서에서 측정된 데이터들 중에 온도가 15도 이상인 데이터들에 대해 시간적 윈도우 또는 공간적 윈도우를 적용하여 대표값을 선택하는 과정일 수 있다.

[0081] 고정 윈도우 기반의 샘플링 프로세스는 데이터 차원에서 서로 중복되지 않고 연속되는 2개 이상의 윈도우에서 대표값을 각각 선택하는 것으로, 예를 들면, 시간 차원에서 제1 시간 구간($t_1 - t_3$)에서 수집한 입력 데이터들 중에서 특정 기준으로 대표값을 선택하고, 제1 시간 구간에 연속한 제2 시간 구간($t_3 - t_5$)에서 수집한 입력 데이터들 중에서 동일한 특정 기준으로 대표값을 선택하는 것일 수 있다.

[0082] 무빙 윈도우 기반의 샘플링 프로세스는 데이터 차원에서 서로 중복되는 2개 이상의 윈도우에서 대표값을 선택하는 것으로, 예를 들면, 시간 차원에서 제1 시간 구간($t_1 - t_3$)에서 수집한 입력 데이터들 중에서 특정 기준으로 대표값을 선택하고, 제1 시간 구간의 일부 구간과 중복되는 제2 시간 구간($t_2 - t_4$)에서 수집한 입력 데이터들 중에서 동일한 특정 기준으로 대표값을 선택하는 것일 수 있다.

[0083] **B. 복합 차원 기반의 샘플링**

[0084] 복합 차원 기반의 샘플링 프로세스는 두 개 이상의 데이터 차원으로 표현되는 데이터에 대해 각 차원에서 독립적으로 단일 차원 샘플링을 수행하는 프로세스이다. 예를 들어, 어떤 지역에 존재하는 센서에서 수집하는 데이터가 온도, 습도, 조도, 소음 중의 적어도 하나를 포함하는 속성으로 구성되어 있고, 상기 센서는 공간적으로 다양한 위치에 존재하고 있으며, 상기 센서에서 측정된 데이터는 주기적으로 수집되거나 사용자의 설정 또는 센서에서 수집하는 데이터의 값에 따라 비주기적으로 수집될 수 있다. 이러한 데이터 수집 환경에서 온도는 모든 센서에 대해 위치에 관계없이 5분으로 정의되는 고정 윈도우 기반의 샘플링을 수행하고, 습도는 특정 위치를 기준으로 7m의 간격으로 정의되는 고정 윈도우 기반의 샘플링을 수행하고, 조도는 습도와 동일한 위치에서 무빙 윈도우 샘플링을 수행하고, 소음은 측정된 소음 데이터가 일정한 기준 값 이상의 데이터만을 선택하는 비주기적 샘플링을 수행할 수 있다.

[0085] 상기 주어진 윈도우에 대해 대표값을 선택하기 위한 기준에는, 사용자에게 의해 미리 정해진 규칙, 윈도우에 포함된 데이터의 통계적 특징 등이 포함될 수 있다. 예를 들어, 사용자는 주어진 윈도우에 포함된 데이터 중에서 특정 기준에서 가장 가까운 위치의 값, 특정 기준에서 가장 먼 위치의 값, 특정 기준에 중앙 위치의 값 등을 선택하도록 규칙을 정할 수 있다.

[0086] 또한 주어진 윈도우에 포함된 전체 데이터 중에서 평균값(average value), 중앙값(medium value), 최대값(maximum value), 최소값(minimum value), 사분위값(quartile value), 표준편차값(standard deviation value), 최빈도값(most frequent value) 등 다양한 통계적 특징으로 정의되는 값들 중의 하나 또는 이러한 값들의 조합이 될 수 있다. 즉, 주어진 윈도우에 포함된 전체 데이터 중에서 평균값과 표준편차값의 두 가지를 대표값으로 선택할 수도 있다.

[0088] **데이터의 축약에 따른 차원변환**

[0089] 차원 변환은 데이터가 표현하는 데이터 차원의 구조를 변경하여 새로운 차원에서 데이터를 표현하는 과정으로, 예를 들면, 주파수 영역 변환(frequency domain transform), 다변량 분석(multivariate analysis), 비선형 차원 축소(nonlinear dimensionality reduction) 등이 포함될 수 있다.

[0090] 푸리에변환(Fourier transform)과 같은 주파수 영역 변환은 시간 차원 또는 공간 차원에서 표현되는 데이터를 주파수 성분으로 분해하여 주파수 차원에서 표현하는 과정으로, 주파수 성분으로 분해된 데이터 중에서 적절한 차단 주파수(cutting frequency)까지만 포함하도록 제한하여 데이터 축약을 달성할 수 있다.

- [0091] 다변량 분석(multivariate analysis)은 다차원 공간에서 표현되는 데이터를 통계적 방식으로 계산해서 동일한 데이터를 표현할 수 있는 새로운 차원을 찾는 과정으로, 상기 새로운 차원으로 정의되는 공간에서 적절한 통계적 기준으로 차원의 수를 제한하여 데이터 축약을 달성할 수 있다. 다변량 분석은, 주성분분석(principal component analysis), 클러스터링(clustering) 등을 예로 들 수 있다.
- [0092] 비선형 차원 축소는 비선형 주성분분석(nonlinear principal component analysis), 미분동형사상 차원 축소(diffeomorphic dimensionality reduction), 곡선 거리 분석(curvilinear distance analysis) 등 다양한 매니폴드 학습(manifold learning)을 이용하여 비선형적으로 차원의 수를 축소하여 데이터 축약을 달성할 수 있다.
- [0094] **데이터의 축약에 따른 샘플링과 차원변환의 결합**
- [0095] 샘플링과 차원변환의 결합은 샘플링과 차원변환을 순차적으로 수행하는 과정으로, 예를 들면, 입력 데이터를 샘플링한 후, 샘플링된 데이터를 차원 변환하거나 입력 데이터의 차원을 변환한 후 변환된 차원에서 입력 데이터를 샘플링하여 데이터의 개수를 줄이는 것일 수 있다.
- [0097] 도 3a 내지 3c는 본 발명의 실시 예에 따른 데이터 축약에서 단일 차원 기반의 샘플링을 설명하기 위한 도면들이다.
- [0098] 도 3a 내지 3c에서는, 시간 차원에서 고정 윈도우를 이용하여 평균을 대표값으로 선택하는 시간 차원 기반의 샘플링의 예를 나타낸 것으로, 도 3a는 그래프 형태의 원본 데이터를 도시한 것이고, 도 3b 및 3c에는 시간 차원 기반의 샘플링에 따라 서로 다른 크기의 고정 윈도우를 이용하여 원본 데이터를 샘플링한 그래프 형태의 축약 데이터를 도시한 것이다.
- [0099] 도 3a에서 시간 차원에서 원본 데이터가 수집되는 시간 간격을 unit1이라고 할 때, 도 3b에 도시한 축약 데이터는 $5 \times \text{unit1}$ 의 시간 간격(unit2)으로 설정된 고정 윈도우를 이용하여 원본 데이터를 샘플링한 것이고, 도 3c는 $10 \times \text{unit1}$ 의 시간 간격(unit3)으로 설정된 고정 윈도우를 이용하여 원본 데이터를 샘플링한 것이다.
- [0101] 도 4는 본 발명의 실시 예에 따른 데이터 축약에서 복합 차원 기반의 샘플링을 설명하기 위한 도면이다.
- [0102] 도 4는 공간 차원과 시간 차원으로 이루어진 복합 차원에서 표현 가능한 원본 데이터의 샘플링을 나타낸 것으로, 참조 번호 41은 서로 다른 장소에 설치된 2개의 센서(sensor1, sensor2)에서 일정 시간 간격으로 수집한 원본 데이터로서, 테이블 형태의 센서 데이터를 나타내고, 참조번호 43은 원본 데이터(41)를 공간 차원에서 축약한 축약 데이터이고, 참조번호 45는 원 데이터(41)를 시간 차원으로 축약한 축약 데이터이다.
- [0103] t11, t12, t13 및 t14는 제1 센서(sensor1)가 Time1, Time2, Time3 및 Time4에서 각각 수집한 온도 데이터이고, t21, t22, t23, t24는 제2 센서(sensor2)가 Time1, Time2, Time3 및 Time4에서 각각 수집한 온도 데이터이다.
- [0104] h11, h12, h13 및 h14는 제1 센서(sensor1)가 Time1, Time2, Time3 및 Time4에서 각각 수집한 습도 데이터이고, h21, h22, h23 및 h24는 제2 센서(sensor2)가 Time1, Time2, Time3 및 Time4에서 각각 수집한 습도 데이터이다.
- [0105] l11, l12, l13, l14는 제1 센서(sensor1)가 Time1, Time2, Time3, Time4)에서 각각 수집한 광도 데이터이고, l21, l22, l23, l24는 제2 센서(sensor2)가 Time1, Time2, Time3 및 Time4에서 각각 수집한 광도 데이터이다.
- [0106] v11, v12, v13 및 v14는 제1 센서(sensor1)가 Time1, Time2, Time3 및 Time4)에서 각각 수집한 전압 데이터이고, v21, v22, v23 및 v24는 센서(sensor2)가 Time1, Time2, Time3 및 Time4)에서 각각 수집한 전압 데이터이다.
- [0107] 이와 같이, 원본 데이터는 서로 다른 장소에 설치된 2개의 센서(sensor1, sensor2)가 일정 시간 간격으로 수집한 데이터이므로, 공간 차원과 시간 차원으로 이루어진 복합 차원으로 표현할 수 있다.
- [0108] 이러한 센서 데이터에 복합 차원 기반의 샘플링 프로세스를 적용하면, 복합 차원에서 표현되는 원본 데이터는 공간 차원에서 표현되는 축약 데이터 및/또는 시간 차원에서 표현되는 축약 데이터로 축약될 수 있다. 예를 들

면, t11과 t21 중에서 어느 하나를 대표값으로 선택하거나 h11과 h21 중에서 어느 하나를 대표값으로 선택하는 것은 복합 차원에서 표현되는 원본 데이터를 공간 차원에서 표현되는 데이터로 축약하는 것이다. t11과 t12 중에서 어느 하나를 대표값으로 선택하거나 h11과 h12 중에서 어느 하나를 대표값으로 선택하는 것은 복합 차원에서 표현되는 데이터를 시간 차원에서 표현되는 데이터로 축약하는 것이다.

- [0110] 도 5는 본 발명의 다른 실시 예에 따른 데이터 축약에서 복합 차원 기반의 샘플링을 설명하기 위한 도면으로서, 소정의 공간에 설치된 센서들의 위치와 의미를 고려한 복합 차원의 데이터 축약을 도식적으로 나타낸 것이다.
- [0111] 도 5에서 사각 박스를 지시하는 참조 번호 51, 53 및 55는 센서들이 설치된 소정의 공간을 지시하고, 각 공간(51, 53, 55) 내에서 원 안에 표기된 숫자들은 센서를 식별하는 번호들이다.
- [0112] 도 5에서는 각 공간에 설치된 센서들이 3가지 케이스로 그룹핑된 상태가 도시된다.
- [0113] CASE1은 공간(51) 내에서 동일한 공간에 설치된 센서들을 그룹화하고, 각 그룹에 속한 센서들에서 측정된 값들 중에서 하나의 대표값을 선택하는 방식으로 데이터를 축약하는 것이다.
- [0114] CASE2는 공간(53) 내에서 동일한 종류의 센서들을 그룹화하고, 각 그룹에 속한 센서들에서 측정된 값들 중에서 하나의 대표값을 선택하는 방식으로 데이터를 축약하는 것이다.
- [0115] CASE3은 특별한 의미를 갖는 기준으로 센서들을 그룹화하고, 각 그룹에 속한 센서들에서 측정된 값들 중에서 하나의 대표값을 선택하는 방식으로 데이터를 축약하는 것이다. CASE3에 센서들을 그룹화하는 기준은 중앙을 기준으로 한 왼쪽 영역과 오른쪽 영역으로 구분한 것이다.
- [0117] 이하, 메타 최적화기에서 설정하는 축약 기준 정보, 학습 기준 정보 및 지식 증강 기준 정보에 대해 상세히 설명한다.
- [0118] 전술한 바와 같이, 메타 최적화기(10)는 입력 데이터의 스키마 정보를 참조하여 축약 기준 정보, 학습 기준 정보, 및 지식 증강 기준 정보를 설정한다.
- [0119] 스키마 정보(schema information)는 입력 데이터와 함께 제공되는 메타 데이터 또는 입력 데이터의 특정 영역에 저장된 메타 데이터를 분석하여 획득하거나 사용자의 입력으로부터 획득할 수 있다.
- [0120] 이러한 스키마 정보(schema information)는 축약 기준 정보, 학습 기준 정보, 지식 증강 기준 정보를 포함하도록 구성될 수 있다. 스키마 정보의 내용은 미리 정해진 규칙에 따라 서술되거나 또는 온톨로지와 같은 구조화된 지식으로 표현되는 지식 사전의 형식으로 서술될 수 있다.
- [0122] **축약 기준 정보**
- [0123] 축약 기준 정보는 데이터 차원에 대한 정보와 데이터 축약에 대한 정보를 포함한다. 데이터 축약에 대한 정보는 주기적 샘플링을 위한 기준 정보, 비주기적 샘플링을 위한 기준 정보, 고정 윈도우 샘플링을 위한 기준 정보, 무빙 윈도우 샘플링을 위한 기준 정보 중의 적어도 하나의 정보를 포함하도록 구성될 수 있으며, 추가로, 샘플링 기준과는 무관하게 적용되는 공통 기준 정보를 더 포함할 수 있다.
- [0124] 주기적 샘플링과 관련된 기준 정보는 데이터 차원에서 윈도우의 위치를 설정하기 위한 윈도우 사이의 간격 정보, 및 대표값을 선택하기 위한 윈도우의 크기 정보를 포함할 수 있다.
- [0125] 비주기적 샘플링과 관련된 기준 정보는 비주기적으로 윈도우를 선택하기 위한 조건 정보, 및 대표값을 선택하기 위한 윈도우의 크기 정보를 포함할 수 있다.
- [0126] 고정 윈도우 샘플링과 관련된 기준 정보는 데이터 차원에서 다수의 윈도우가 서로 중복되지 않게 주어지는 윈도우의 크기 정보를 포함할 수 있다.
- [0127] 무빙 윈도우 샘플링과 관련된 기준 정보는 데이터 차원에서 서로 중복되는 윈도우의 위치를 설정하기 위한 간격 정보 및 대표값을 선택하기 위한 윈도우의 크기 정보를 포함할 수 있다.
- [0128] 상기 샘플링 기준과는 무관하게 적용되는 공통 기준 정보는 윈도우의 크기에서 대표값을 선택하기 위한 기준 정

보를 포함할 수 있다.

[0130] **학습 기준 정보**

[0131] 본 발명에서는 학습 모델의 성능 또는 학습 결과의 신뢰도(또는 정확성)는 데이터 축약의 적절성을 평가하기 위한 지표로 사용된다.

[0132] 학습 기준 정보는 축약 데이터로 학습 모델을 학습하는 과정에서, 학습의 반복을 제한하는 조기 종료 조건(early stop condition), 수렴 추이 조건(convergence trend condition) 등을 포함할 수 있으며, 추가로, 학습의 성능을 계산하기 위한 학습 신뢰도 조건(learning reliability condition) 등을 더 포함할 수 있다.

[0133] 학습 신뢰도 조건은 학습 성능 평가뿐만 아니라 학습의 반복을 제한하는 조건으로 사용할 수도 있다.

[0134] 학습 모델의 특성에 따라 달라질 수 있는 학습 기준의 선택은 스키마 정보에 의해 결정되기 때문에, 학습 기준은 다양하게 구성할 수 있다. 따라서, 본 발명에서는 학습 기준을 한정하지 않는다.

[0135] 학습의 대상이 되는 데이터, 즉, 학습 데이터는 훈련 데이터(train dataset), 검증 데이터(validation dataset), 테스트 데이터(test dataset)를 포함할 수 있다.

[0136] 훈련 데이터는 학습 모델을 훈련하는 데 사용할 수 있다. 검증 데이터는 적절한 데이터 축약을 선택하는 데 사용할 수 있다. 테스트 데이터는 선택된 데이터 축약의 유효성 또는 타당성을 판단하는 데 사용할 수 있다. 훈련 데이터와 검증 데이터는 동일한 데이터일 수 있다.

[0137] 조기 종료 조건(early stop condition)과 수렴 추이 조건(convergence trend condition)은 학습 반복을 통해 학습 모델을 최적화하는 학습 과정에서 암기 효과(memorization effect)를 방지하기 위해 사용하는 정규화(regularization)의 한 형태로, 학습 결과가 미리 지정한 학습 신뢰도 조건을 만족하기 이전에 수행할 수 있는 반복 학습의 범위를 제한한다.

[0138] 상기 학습 신뢰도는 분류 모델(classification model)에서 주로 사용되는 정밀도(precision), 정확도(accuracy), AUC(area under curve) 등의 지표, 회기 모델(regression model)에서 주로 사용되는 RMSE(root mean squared error), MAE(mean absolute error), RAE(relative absolute error), RSE(relative square error), 결정 계수(coefficient of determination) 등의 지표, 클러스터링 모델(clustering model)에서 주로 사용하는 클러스터의 결집도(compactness of a cluster), 클러스터 중심까지의 최대 거리(maximal distance to cluster center), 클러스터 사이의 거리(distance between clusters) 등의 지표를 이용하거나 조합하여 다양한 방식으로 계산할 수 있다.

[0139] 상기 데이터 축약의 적절성은 학습 과정이나 학습 결과가 학습 기준에서 정한 조건을 만족하는지 여부를 통해 평가할 수 있다. 조기 종료 조건이나 수렴 추이 조건은 학습 반복을 제한하기 위해 사용하므로, 학습 결과나 학습 과정이 미리 정한 학습 신뢰도 조건을 만족하지 못한 상태에서 학습 결과나 학습 과정이 조기 종료 조건이나 수렴 추이 조건을 만족하는 상황이 발생하면, 학습 과정은 자동으로 종료된다.

[0140] 학습이 종료되는 경우, 데이터 축약은 적절하지 못한 것으로 판정할 수 있으며, 적절한 데이터 축약이 가능하도록 축약 기준 정보의 변동을 토대로 반복 학습이 수행될 수 있다.

[0141] 학습의 반복이 조기 종료 조건(early stop condition)이나 수렴 추이 조건(convergence trend condition)을 만족하지는 않지만, 학습 신뢰도 조건을 만족하면, 학습 과정은 자동으로 종료된다. 이와 같은 상황에서 학습 과정이 종료되는 경우, 데이터 축약은 적절한 것으로 판정할 수 있다, 학습 결과는 학습 이력으로 저장된다.

[0142] 저장되는 학습 이력은 자가 학습 과정에서 발생하는 정보들, 예를 들면, 입력 데이터, 스키마 정보, 축약 기준 정보, 축약 데이터 정보, 학습 기준 정보, 학습 데이터 정보, 학습 모델 정보, 학습 결과 정보, 지식 증강 기준 정보 등을 포함할 수 있다.

[0143] 데이터 축약이 적절한 것으로 판정되고 지식 증강 기준을 만족하면, 축약 기준 정보를 최적화하는 지식 증강 프로세스가 수행된다.

[0145] **지식 증강 기준 정보**

- [0146] 본 발명에서 지식 증강 기준 정보는 축약 기준 정보를 갱신하기 위한 기준 및 조건을 정의한다.
- [0147] 지식 증강 기준 정보는 학습 기준의 제한(또는 반복 학습 기준), 축약 기준의 변동, 이력 누적 기준 등을 포함할 수 있다. 지식 증강 기준 정보가 축약 기준의 변동 정보와 반복 학습 기준 정보를 반드시 포함하는 것은 아니며, 필요한 경우에는 이력 누적 기준 정보만을 포함하도록 구성될 수 있다.
- [0148] 상기 반복 학습 기준 정보는 데이터 축약 기준을 최적화하기 위한 지식 증강 과정에서 만족하여야 하는 학습 기준의 요소를 명시한 것이다.
- [0149] 상기 축약 기준의 변동 정보는 축약 기준을 변동할 수 있는 요소와 범위를 명시한 것이다.
- [0150] 상기 이력 누적 기준은 축약 기준 정보를 최적화하기 위한 지식 증강을 수행하기에 앞서 만족해야 하는 조건을 명시한 것으로, 학습 이력 누적 조건, 축약 기준의 변동 조건 등을 포함할 수 있으며, 이러한 조건을 만족하지 않는 경우에는 축약 기준 정보를 최적화하기 위한 지식 증강이 수행되지 않는다.
- [0152] 도 6a는 본 발명의 일 실시 예에 따른 스키마 정보에 포함된 축약 기준 정보의 데이터 구조를 나타낸 도면이다.
- [0153] 도 6a를 참조하면, 축약 기준 정보의 데이터 구조는, 예를 들면, 5개의 필드들(F1~F5)을 포함할 수 있다. 제1 필드(F1)에는 DR-ID와 같은 축약 기준 정보의 식별자가 기록될 수 있다. 제2 필드(F2)에는 데이터 차원을 나타내는 정보가 기록될 수 있다. 제3 필드(F3)에는 데이터 축약에 사용되는 윈도우의 종류를 나타내는 정보가 기록될 수 있다. 제4 필드(F4)에는 윈도우의 크기를 나타내는 정보가 기록될 수 있다. 제5 필드(F5)에는 대표값 선택을 위한 기준을 나타내는 정보가 기록될 수 있다. 대표값 선택 기준은 대표값의 속성, 대표값의 종류, 대표값 선택 방법 또는 대표값 계산 방법과 관련된 정보일 수 있다. 필드의 순서는 설계에 따라 다양하게 변경할 수 있다.
- [0154] 제1 필드(F1)에 'DR001', 제2 필드(F2)에 '시간', 제3 필드(F3)에 '고정 윈도우', 제4 필드(F4)에 '10분' 및 제5 필드(F5)에 '평균'이 기록된 경우, 축약 기준 정보는 DR001로 식별되고, 시간 차원에서 윈도우 크기가 10분인 고정 윈도우를 이용하여 선택한 평균값을 대표값으로 선택하는 축약 규칙을 정의한다.
- [0156] 도 6b는 본 발명의 일 실시 예에 따른 스키마 정보에 포함된 학습 기준 정보의 데이터 구조를 나타낸 도면이다.
- [0157] 도 6b를 참조하면, 학습 기준 정보의 데이터 구조는 5개의 필드를 포함할 수 있다. 제1 필드(F1)에는 'LC-ID'와 같은 학습 기준 정보의 식별자(LC-ID, Learning Condition-Identifler)가 기록된다. 제2 필드(F2)에는 학습 신뢰도 계산을 위해 사용되는 데이터의 종류와 관련된 정보가 기록된다. 제3 필드(F3)에는 학습 신뢰도 조건과 관련된 정보가 기록될 수 있다. 제4 필드(F4)에는 학습 신뢰도를 계산하는 기준과 관련된 정보가 기록될 수 있다. 여기서, 학습 신뢰도를 계산하는 기준은 학습 신뢰도의 계산 방법과 관련된 정보일 수 있다. 제5 필드(F5)에는 학습의 조기 종료 조건(early stop condition)과 관련된 정보가 기록될 수 있다.
- [0158] 제1 필드에 'LC001', 제2 필드에 '검증 데이터' 제3 필드에 '5% 이하' 제4 필드에 '평균 제곱근 오차(Root Mean Square Error; RMSE)', 제5 필드에 '2,000회 이상'이 기록된 경우, 학습 기준 정보는 'LC001'로 식별되고, 검증 데이터를 이용하여 학습 신뢰도를 계산하고, 학습 과정에서 학습 신뢰도의 RMSE가 5%이하이거나, 학습의 반복 횟수가 2,000회 이상인 경우 학습을 중단하는 규칙을 정의한다.
- [0159] 다르게, 위의 예에서, 학습 기준 정보는 학습 과정에서 학습의 반복 횟수가 2,000회 미만이고, 검증 데이터를 이용하여 계산한 학습 신뢰도의 RMSE 값이 5% 보다 작은 값에 도달하면 학습 기준을 만족한 것으로 판단하는 규칙으로 정의할 수도 있다.
- [0160] 다르게, 위의 예에서, 학습 기준 정보는 학습의 반복 횟수가 2,000회를 넘는 순간에 RMSE 값이 5% 이상이면 학습 신뢰도가 학습 기준을 만족하지 못한 것으로 판단하는 규칙으로 정의할 수 있다.
- [0162] 도 6c는 본 발명의 일 실시 예에 따른 스키마 정보에 포함된 지식 증강 기준 정보의 데이터 구조를 나타낸 도면이다.
- [0163] 도 6c를 참조하면, 지식 증강 기준 정보는 반복 학습 기준 정보(61), 축약 기준 변동 정보(63) 및 이력 누적 기

준 정보(65)를 포함할 수 있다.

[0164] **반복 학습 기준 정보(61)**

[0165] 반복 학습 기준 정보(61)의 데이터 구조는 3개의 필드들(F1~F3)을 포함할 수 있다. 제1 필드(F1)에는 'KA-ID1'과 같은 반복 학습 기준 정보의 식별자(Knowledge Augmentation IDentifier)가 기록되고, 제2 필드(F2)에는 제한하고자 하는 학습 기준 정보의 식별자(LC-ID)가 기록되고, 제3 필드(F3)에는 축약 기준의 변동횟수가 기록될 수 있다.

[0166] 축약 기준 변동에 따른 학습의 반복횟수가 5회 이하인 경우, 반복 학습 기준 정보(61)는 LC-ID로 식별되는 학습 기준 정보에서 제한하는 조건(예를 들면, 학습의 반복횟수가 2000회 이하이고, RMSE가 5%미만인 조건)을 만족하지 못하면, 축약 기준을 변동하여 반복 학습을 수행할 수 있지만, 이러한 축약 기준 변동 횟수가 5회까지만 허용하는 규칙을 정의한다. 즉, 반복 학습 기준 정보(61)에서 정의하는 규칙은 축약 기준을 5회 변동하는 과정에서 학습 결과가 학습 기준 정보에서 제한하는 조건이 만족되면, 그 학습 결과를 학습 이력으로서 저장하고, 축약 기준의 변동을 종료하지만, 축약 기준을 5회 변동하기까지 학습 결과가 학습 기준 정보에서 제한하는 조건을 만족하지 못하면, 그 학습 결과를 학습 이력으로 저장하는 않음을 규정한 것이다. 여기서, 저장되는 학습 이력은 자가 학습 과정에서 발생하는 정보들, 예를 들면, 입력 데이터 정보, 스키마 정보, 학습 모델 정보, 축약 기준 정보, 학습 기준 정보, 학습 데이터 정보, 학습 모델 정보 학습 결과 정보, 지식 증강 기준 정보 등을 포함할 수 있다.

[0167] **축약 기준 변동 정보(63)**

[0168] 축약 기준 변동 정보(63)의 데이터 구조는 5개의 필드들(F1~F5)을 포함할 수 있다. 제1 필드(F1)에는 변동 대상에 해당하는 축약 기준 정보의 식별자(DR-ID)가 기록될 수 있고, 제2 필드(F2)에는 상기 식별자(DR-ID)에 의해 식별되는 축약 기준 정보 내에서 변동되는 변동 요소와 관련된 정보가 기록될 수 있고, 제3 필드(F3)에는 제2 필드(F2)에 기록된 변동 요소의 변동 범위와 관련된 정보가 기록될 수 있고, 제4 필드(F4)에는 상기 변동 범위 내에서 특정되는 변동 기준과 관련된 정보가 기록될 수 있고, 제5 필드(F5)에는 상기 변동 기준을 임의로 변동시키는 규칙과 관련된 정보가 기록될 수 있다.

[0169] 예를 들면, 변동 요소가 고정 윈도우의 크기이고, 변동 범위가 0.5배, 1.0배 및 1.5배를 포함하고, 변동 기준이 10분이고, 임의성 규칙이 10분의 30.0%인 경우, 축약 기준 변동 정보(63)는 10분의 고정 윈도우의 크기를 5분, 10분, 15분의 고정 윈도우로 확장 또는 축소하고, 10분의 30% 범위 내에서 고정 윈도우의 크기를 임의적으로 변경하는 축약 기준의 변동을 규정한 것이다.

[0170] 고정 윈도우의 크기를 임의적으로 변경을 위해, 다양한 윈도우 설정을 위해 랜덤 함수를 이용하거나 교배 및 돌연변이 과정을 통해 임의성을 발생시키는 유전자 알고리즘이 활용될 수 있다.

[0171] 이에 따라, 윈도우의 크기를 [3분, 10분, 17분], [7분, 13분, 15분], [5분, 9분, 16분] 등으로 다양하게 자동 설정할 수 있다.

[0172] **이력 누적 기준 정보(65)**

[0173] 반복 학습 기준의 규칙에 따른 수행이 완료되면, 이어, 이력 누적 기준의 규칙에 따른 수행이 시작된다.

[0174] 이력 누적 기준 정보(65)는 학습 이력의 누적 기준을 정의하는 규칙으로, 학습 누적과 지식 증강 개시를 위한 축약 기준 변동을 정의한다.

[0175] 이력 누적 기준 정보(65)의 데이터 구조는 3개의 필드들(F1~F3)을 포함할 수 있다. 제1 필드(F1)에는 'KA-ID2'과 같은 이력 누적 기준 정보의 식별자가 기록될 수 있고, 제2 필드(F2)에는 학습 이력의 누적 횟수와 관련된 정보가 기록될 수 있고, 제3 필드(F3)에는 지식 증강을 수행하기 위한 축약 기준의 변동 횟수가 기록될 수 있다.

[0176] 학습 결과를 이력으로 저장하기 위한 누적 횟수가 15회 이상이고, 지식 증강을 수하기 위한 축약 기준의 변동 횟수가 6회 이상이면, 학습 이력이 저장될 때마다 축약 기준 정보를 최적화하는 지식 증강이 수행된다. 그러나, 학습 이력 누적 또는 축약 기준 변동 중 하나라도 만족하지 않으면, 지식 증강은 수행되지 않는다.

[0178] 도 7은 본 발명에 일 실시 예에 따른 스키마 정보를 온톨로지로 표현한 것이다.

- [0179] 도 7에 도시한 온톨로지는 축약 기준 정보를 온톨로지로 표현한 것이다. 본 발명에서 예시한 규칙 또는 구조화된 지식은 다양한 방식으로 설정이 가능하며, 본 발명에서 제시한 예시로 한정하지는 않는다.
- [0181] 도 8은 본 발명의 제2 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 장치를 도시한 블록도이다.
- [0182] 도 8을 참조하면, 본 발명의 제2 실시 예에 따른 데이터 메타 스케일링 장치는, 메타 최적화기(10)와, 축약기(20)와, 학습기(30), 평가기(40) 및 메타 정보 저장소(50)를 포함한다.
- [0183] 상기 메타 정보 저장소(50)는 학습 이력 정보를 저장하며, 학습 이력 정보는, 자가 학습 과정에서 발생하는 정보들, 즉, 상기 메타 최적화기(10)와 축약기(20)와 학습기(30) 및 평가기(40)에서 입출력되는 모든 정보를 포함하며, 예를 들면, 입력 데이터 정보, 스키마 정보, 학습 모델 정보, 축약 기준 정보, 축약 데이터 정보, 학습 기준 정보, 학습 데이터 정보, 학습 모델 정보, 학습 결과 정보, 지식 증강 기준 정보 등을 포함한다.
- [0184] 한편, 상기 메타 최적화기(10)와 축약기(20)와 학습기(30) 및 평가기(40)는, 상호 연동을 위해 학습 이력 정보를 입출력하는 과정에서 상기 메타 정보 저장소(50)를 활용할 수 있다. 예를 들어, 메타 최적화기(10)가 스키마 정보에서 추출하거나 사용자 입력에 따라 제공된 축약 기준 정보, 학습 기준 정보, 지식 증강 기준 정보를 메타 정보 저장소(50)에 저장하고, 이어서 축약기(20)에 메타 정보 저장소(50)의 저장 위치에 대한 정보를 전달하면, 축약기(20)는 상기 축약 기준 정보를 상기 메타 정보 저장소(50)로부터 읽어 들여, 입력된 데이터의 차원을 상기 축약 기준 정보에 따라 축약할 수 있다.
- [0185] 또한, 상기 축약기(20)가 축약 데이터를 상기 메타 정보 저장소(50)에 저장하면, 학습기(30)가 상기 저장된 축약 데이터를 메타 정보 저장소(50)에서 읽어와 학습 데이터로 구성하여 기계 학습을 수행할 수 있다.
- [0186] 마찬가지로, 학습기(30)도 학습 결과 정보를 상기 메타 정보 저장소(50)에 저장하면, 평가기(40)는 상기 메타 정보 저장소(50)로부터 학습 결과 정보를 읽어들이, 학습 결과가 학습 기준을 만족하는지 여부를 판단할 수 있다.
- [0187] 최종적으로, 메타 최적화기(10)는 평가기(40)의 판단 결과를 바탕으로 축약 기준 정보의 갱신 또는 지식 증강을 수행할 수 있다.
- [0188] 상술한 제2 실시 예에 따르면, 데이터 메타 스케일링 장치는, 학습 이력 정보를 누적하고, 누적된 학습 이력 정보를 저장하고, 지식 증강 기준을 만족하도록 학습 이력 정보가 충분히 저장되면, 학습 이력을 분석하여 최적의 축약 기준을 찾아 스키마 정보를 자동 갱신한다. 이러한 과정을 통해 자가 학습을 구성하는 절차를 자동화할 수 있으며, 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있다.
- [0190] 도 9는 본 발명의 제3 실시 예에 따른 자가 학습을 위한 데이터 메타 스케일링 장치를 도시한 블록도이다.
- [0191] 도 9를 참조하면, 본 발명의 제3 실시 예에 따른 데이터 메타 스케일링 장치는, 메타 최적화기(100), 다수의 축약기(200: 1, 2, ..., N), 다수의 학습기(300: 1, 2, ..., M), 평가기(400) 및 메타 정보 저장소(500)를 포함한다.
- [0192] 본 발명의 제3 실시 예에 따른 데이터 메타 스케일링 장치는, 하나의 축약기와 하나의 학습기를 구비하는 도 1 및 8의 실시 예들과는 다르게 다수의 축약기와 다수의 학습기를 구비함으로써, 다수의 학습기가 다수의 축약기(200)에 의해 축약된 다수의 데이터의 학습을 병렬로 처리할 수 있다.
- [0193] 이때, 상기 메타 최적화기(100)는 다수의 축약기(200)에 각각 제공하는 상기 다수의 축약 기준 정보를 설정하기 위하여 복합 차원의 데이터 축약기(110)를 포함할 수 있다.
- [0194] 상기 복합 차원의 데이터 축약기(110)는 데이터의 속성을 표현할 수 있는 다양한 차원에서 정의되는 다양한 축약 단위가 조합된 다수의 축약 기준 정보를 포함하는 축약 기준 정보 집합을 설정한다.
- [0195] 구체적으로, 상기 복합 차원의 데이터 축약기(110)는 데이터를 표현할 수 있는 다양한 차원들의 축약 단위들을 유전자 알고리즘을 이용하여 조합하여 축약 기준 정보 집합(축약 기준 정보 1 내지 축약 기준 정보 N)을 설정한다.
- [0196] 상기 축약 기준 정보 1 내지 축약 기준 정보 N은 다수의 축약기(200)로 제공되고, 상기 다수의 축약기(200) 각

각은, 각자의 축약 기준 정보에 따라서 입력된 데이터를 축약한다. 여기서, 다수의 축약기(200)로 입력된 데이터는 동일하나, 적용된 축약 기준 정보가 다르므로 상기 다수의 축약기(20)로부터 출력되는 축약 데이터들은 서로 상이하게 된다.

- [0197] 서로 다른 축약 기준 정보에 따라서 축약된 축약 데이터들은 다수의 학습기(300)로 각각 제공된다. 상기 다수의 학습기(300)는 서로 다른 학습 모델로 구성되며, 서로 다른 다수의 축약 기준 정보들에 따라서 축약된 축약 데이터를 학습한다. 즉, 다수의 학습기(1, 2, ..., M)가 상기 축약 기준 정보 1에 따라 축약된 축약 데이터에 대해 병렬 학습을 수행하고, 이러한 병렬 학습은 다수의 학습기(1, 2, ..., M)가 축약 기준 정보 N에 따라 축약된 축약 데이터(M)의 병렬 학습을 완료할 때까지 진행된다. 따라서, 다수의 학습기(1, 2, ..., M)는 N*M 개의 학습 결과를 평가기(400)로 제공한다.
- [0198] 한편, 다수의 학습기(1, 2, ..., M)는 하나의 공통된 학습 기준 정보를 기반으로 서로 다른 다수의 축약 기준 정보들에 따라서 축약된 다수의 축약 데이터들에 대해 학습을 병렬적으로 수행할 수 있지만, 축약 데이터별로 서로 다른 학습 기준 정보를 기반으로 학습을 병렬적으로 수행할 수도 있다. 이 경우, 메타 최적화기(100)는 서로 다른 학습 기준 정보를 설정하도록 구성될 수 있다.
- [0199] 평가기(400)는 N*M 개의 학습 결과에 대한 학습 신뢰도가 학습 기준을 만족하는지 여부를 판단한다. 이때, 학습 모델과 축약 데이터의 다양한 조합에 의해 학습 결과의 신뢰도는 서로 상이한 값을 가지고 학습 모델의 특성(예를 들어, 초매개변수)도 서로 상이할 수 있다.
- [0200] 평가기(400)는 상기 다수의 학습기(300)에서 제공되는 학습 결과의 학습 신뢰도가 학습 기준을 만족하는지를 판단하고, 메타 최적화기(100)는 평가기(40)의 판단 결과를 기반으로 다수의 축약 기준 정보 전체 또는 일부를 갱신한다.
- [0201] 학습 결과의 학습 신뢰도가 학습 기준을 만족하지 못하는 경우, 메타 최적화기(100)는 지식 증강 기준 정보를 기반으로 축약 기준 정보를 갱신한다. 학습 결과의 학습 신뢰도가 학습 기준을 만족하는 경우, 메타 최적화기(100)는 학습 결과를 학습 이력으로 자동 저장하는 과정을 통해 지식 증강 프로세스를 시작한다.
- [0202] 메타 최적화기(100)는 지식 증강 기준을 만족하도록 학습 이력이 충분히 저장된 후 학습 이력을 분석하여 축약 기준을 최적화하는 프로세스를 수행한다. 이러한 과정을 통해 자가 학습을 구성하는 절차를 자동화할 수 있으며, 지속적인 지식 증강을 통해 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있다.
- [0204] 도 10은 도 1에 도시한 데이터 메타 스케일링 장치를 교통 정보 예측 시나리오에 적용하는 예를 설명하기 위한 도면이다.
- [0205] 도 10을 참조하면, 교통 정보 예측 시나리오에 적용될 수 있는 축약 기준 정보의 예는 시간으로 정의되는 데이터 차원, 고정 윈도우로 정의되는 윈도우 종류, 10분의 정의되는 윈도우 크기, 평균으로 정의되는 대표값 선택 기준을 포함할 수 있다. 이러한 축약 기준 정보는 시간 차원에서 윈도우 크기 10분의 고정 윈도우를 대상으로 평균을 구한 결과를 대표값으로 선택하여 교통 데이터를 축약하는 규칙을 의미한다.
- [0206] 교통 정보 예측 시나리오에 적용될 수 있는 학습 기준 정보의 예는 검증 데이터로 정의되는 데이터 종류, 0.15% 이하로 정의되는 학습 신뢰도 조건, RMSE로 정의되는 학습 신뢰도 계산 기준, 2000회 이상으로 정의되는 조기 종료 조건을 포함할 수 있다. 이러한 학습 기준 정보는 교통 예측 모델에 대해 검증 데이터를 이용하여 학습 신뢰도를 계산하고, 학습 과정에서 학습 신뢰도의 RMSE가 0.15%보다 이하이거나, 학습의 반복 횟수가 2,000회를 넘는 경우 학습을 중단하는 규칙을 의미한다.
- [0207] 교통 정보 예측 시나리오에 적용되는 지식 증강 기준 정보는 5회 이내의 축약 기준의 변동 횟수, 윈도우 크기로 정의되는 변동 요소, 5분, 10분, 15분으로 정의되는 변동 범위, 15회 이상으로 정의되는 학습 누적 횟수, 6회 이상의 축약 기준 변동 횟수로 정의되는 지식 증강 개시 조건을 포함할 수 있다. 이러한 지식 증강 기준 정보는 축약 기준 정보의 변동을 기반으로 하는 학습을 5회 이내에서 반복하고, 고정 윈도우 크기는 [5분, 10분, 15분]의 세 종류를 설정하고, 학습 결과를 이력으로 저장한 누적 횟수가 15회 이상이고, 축약 기준 변동의 횟수가 6회 이상이면, 학습 결과를 이력으로 저장할 때마다 축약 기준 정보를 최적화하는 지식 증강이 수행되는 규칙을 의미한다.
- [0208] 메타 최적화기(10)는 교통 정보 예측 시나리오에 적용되는 축약 기준 정보를 축약기(20)에 제공한다. 축약기(20)는 시간 차원에서 5분, 10분 및 15분의 윈도우를 이용하여 대표값을 선택하는 축약 프로세스를 수행한다.

학습기(30)는 축약기(20)에 의해 축약된 데이터에 대해 학습을 수행한다. 평가기(40)는 학습기(30)의 학습 결과가 학습 기준 정보에서 규정한 학습 기준을 만족하는지를 판단하게 된다. 예를 들어, 10분 단위의 축약에서 학습 신뢰도의 RMSE가 0.13%인 경우, RMSE가 0.15%보다 작은 규칙을 만족하므로, 그 학습 결과는 이력으로 저장되고, 지식 증가 기준 정보의 규칙에 따른 수행은 완료된다.

[0209] 한편, 교통 정보 예측 시나리오에 적용하는 스키마 정보는 데이터 차원이 공간 차원이나 의미 차원인 경우의 축약 기준 정보를 포함할 수 있다. 예를 들어, 공간 차원의 축약 기준 정보와 관련해, 축약기(20)는 통행 속도가 측정된 도로 지점이 속한 행정구역(예를 들어, 시/군/구) 또는 용도지역(예를 들어, 주거지역, 중심상업지역 등)과 같은 공간 단위로 교통 데이터를 축약하고, 공간 단위로 축약된 축약 데이터를 이용하여 예측 모델을 산출할 수도 있다.

[0210] 구체적으로, 메타 최적화기(10)는 인접한 도로의 교통량을 고려하기 위해, 특정 블록 안에 있는 도로에서 측정된 차량의 속도 데이터들에 대한 축약 기준 정보를 설정할 수 있다. 이 경우, 특정 지점의 통행 속도를 예측하기 위해, 특정 지점이 속한 행정 구역의 교통량을 측정된 데이터뿐만 아니라 인접한 행정 구역의 교통량을 측정된 데이터도 추가로 활용할 수도 있다. 이 경우, 축약 기준 정보는 (데이터 차원: 공간), (윈도우 종류: 고정 윈도우), (윈도우 크기: 3블록), (대표값 선택 기준: 평균 속도)의 규칙을 설정할 수 있다. 이 규칙은 공간 차원에서 3블록의 고정 윈도우를 이용하여 평균 속도를 대표값으로 선택하는 데이터 축약 프로세스를 의미한다.

[0211] 또한, 메타 최적화기(10)는 의미 정보와 시간 정보를 결합한 축약 기준 정보를 설정할 수 있다. 이 경우, 축약 기준 정보는 (데이터 차원: 공간), (축약 위치: 종로구), (윈도우 크기: 상업 지구), (데이터 차원: 시간), (축약 범위: 08:00~09:30), (윈도우의 종류: 고정 윈도우), (윈도우 크기: 10분), (대표값 선택 기준: 평균 속도)를 포함할 수 있다. 이 규칙은 종로구에 속한 상업 지구라는 의미 차원으로 정의되는 공간에서 08:00~09:30의 시간 윈도우에 대해 10분의 고정 윈도우로 평균 속도를 대표값으로 선택하는 데이터 축약 프로세스를 의미한다.

[0213] 도 1에 도시한 데이터 메타 스케일링 장치의 다른 적용 예로, 도 1에 도시한 데이터 메타 스케일링 장치는 전력 수요 예측 서비스에 적용할 수도 있다.

[0214] 축약 기준을 적절히 설정하여 에너지 사용량의 결측값 제거(missing value removal) 및 노이즈 제거(noise removal)를 통한 양질의 에너지 사용량 데이터를 생성할 수 있다.

[0215] 에너지 수요 관리를 위해서 전력 에너지를 소비하는 냉난방 장치 및 조명기기들의 전력 사용량 데이터가 일정 시간 간격마다 측정하여, 향후 특정 시점의 에너지 수요 예측을 위한 정확한 학습 모델을 생성할 필요가 있다. 이때, 개별 기기로부터 측정된 전력 사용량은 기상 변화 및 특정 행사 개최 등의 외부 요인에 의해서 불규칙한 사용 패턴을 보이는 경우가 많을 뿐만 아니라, 장비 이상 및 사용자의 데이터 공개 거부 등으로 인해 결측값이 존재할 수 있다.

[0216] 이에, 본 발명의 데이터 축약을 이용할 경우, 데이터 축약 단위 변환을 통해 측정 데이터의 일부 결측값 제거 및 노이즈 제거 또한 가능하다.

[0217] 예를 들어, 축약 기준 정보가 (데이터 차원: 공간), (축약 위치: 연구동), (윈도우 크기: 3층), (데이터 차원: 시간), (축약 범위: 08:00~19:00), (윈도우 종류: 고정 윈도우), (윈도우 크기: 10분), (대표값 선택 기준: 최대 전력 사용량)을 포함할 때, 이 축약 기준 정보는 연구동(research building)의 3층이라는 의미 차원으로 정의되는 공간에서 08:00~19:00의 시간 윈도우에 대해 10분의 고정 윈도우로 정해진 범위 내에서 최대 전력 사용량을 대표값으로 선택하는 데이터 축약 프로세스를 의미한다.

[0218] 메타 최적화기(10)는 전력 수요 예측 서비스에 적용되는 축약 기준 정보를 축약기(20)에 제공하고, 축약기(20)는 상기 축약 기준 정보를 토대로 데이터 축약을 실시하고, 학습기(30)는 주어진 전력 수요 예측 모델에 대한 학습을 수행하고, 평가기(40)는 학습 결과 정보가 학습 기준을 만족하는지 여부를 판단하게 된다. 이때, 학습 결과 정보에 따른 학습 결과가 학습 기준을 만족하면 학습 결과는 이력으로 저장되고, 지식 증가 기준 정보에 따른 수행은 완료된다.

[0220] 도 1에 도시한 데이터 메타 스케일링 장치의 또 다른 적용 예로, 도 1에 도시한 데이터 메타 스케일링 장치는 풍력발전시스템의 발전 효율 최적화에 적용될 수 있다.

- [0221] 이 적용 예에서는, 풍향과 풍속의 변화에 따라 풍력발전기의 블레이드 날개의 각도 조절 타이밍을 최적화할 수 있도록 발전량 데이터를 저장하기 위한 적절한 축약 기준을 설정할 필요가 있다. 이때, 풍향과 풍속은 미기상 바람 예측 모델(micro-meteorological wind prediction model)을 이용하여 예측할 수 있다. 미기상 바람 예측 모델은 수치 예측 모델(numerical prediction model), 기계 학습 예측 모델(machine learning prediction model), 그리고 수치 예측 모델과 기계 학습 예측 모델을 융합한 하이브리드 모델(hybrid model) 등 다양한 모델을 적용할 수 있다.
- [0222] 예측되는 풍향과 풍속의 변화에 따른 블레이드 날개의 각도 조절을 위해, 다양한 전략 및 모델이 존재하고 있으며, 본 발명에서는 상기 전략과 모델을 제한하지 않는다.
- [0223] 메타 스케일링 장치는 풍력발전시스템의 발전 효율 최적화에 적용되는 예에서, 메타 최적화기(10)는 풍력발전량과 관련된 축약 기준 정보를 축약기(20)에 제공하고, 축약기(20)는 상기 축약 기준 정보를 토대로 데이터 축약을 실시하고, 학습기(30)는 축약된 데이터를 이용하여 주어진 풍력발전량 예측 모델에 대한 학습을 수행하고, 평가기(40)는 학습기(30)의 학습 결과가 학습 기준을 만족하는지 여부를 판단하게 된다. 이때, 학습 결과가 학습 기준을 만족하면 학습 결과는 이력으로 저장되고, 지식 증강 기준 정보의 규칙에 따른 수행은 완료된다.
- [0225] 본 발명에서는 지식 증강 기준 정보에 따른 규칙을 토대로 학습 이력을 누적하여 저장하고, 지식 증강 기준 정보에 따른 규칙을 만족하도록 학습 이력이 충분히 저장되면, 학습 이력을 분석하여 축약 기준을 최적화하고, 스키마 정보에 최적화된 축약 기준 정보를 추가하여 스키마 정보를 자동으로 갱신하는 과정을 통해 자가 학습을 실현한다.
- [0226] 이하에서는 스키마 정보를 갱신하기 위해, 최적의 축약 기준을 찾는 과정에 대해 설명하기로 한다.
- [0227] 도 11a 내지 11c는 본 발명의 일 실시 예에 따른 최적의 축약 기준을 찾는 지식 증강 프로세스를 도식적으로 나타낸 도면으로, 도 11a는 하나의 데이터 차원에서 다양한 윈도우 크기를 기반으로 학습기가 학습한 이력을 저장한 결과를 2차원 형태로 나타낸 것이고, 도 11b는 두 개의 데이터 차원에서 다양한 윈도우 크기를 기반으로 학습기가 학습한 이력을 저장한 결과를 3차원 형태로 나타낸 것이다. 그리고 도 11c는 저장된 학습 이력을 이용하여 최적의 윈도우 크기를 찾아 축약 기준 정보를 최적화하는 과정을 나타낸 것이다.
- [0228] 도 11a에서는, 수평축과 수직축으로 정의되는 평면에서 다양한 크기를 갖는 다수의 원이 나타나며, 각 원은 학습 결과의 신뢰도를 의미한다. 여기서, 학습 결과는 주기적으로 반복되는 사건(event)에 대한 센싱 데이터를 학습한 결과이다.
- [0229] 학습 결과의 신뢰도는 원의 크기와 관련된다. 예를 들면, 원의 크기가 클수록 학습의 신뢰도(또는 정확도)가 높다.
- [0230] 각 원의 중심은 수평축 상에서 주기에 따른 상대적인 위치로 표현되고, 수직축 상에서 축약 기준 정보에 따른 윈도우 크기에 따른 위치로 표현된다. 즉, 수평 축은 임의의 데이터 차원에서 반복되는 사건의 센싱 주기에 따라 수집되는 센싱 값을 표현한 것으로, 수평축의 범위는 'D10'로 표시되는 최소값과 'D20'으로 표시되는 최대값으로 정의된다.
- [0231] 수직축은 축약 기준 정보에 따라 데이터 축약 프로세스에서 사용하는 윈도우 크기를 표현한 것으로, 수직축의 범위는 '0'으로 표시되는 최소값과 '50'으로 표시되는 최대값으로 정의된다.
- [0232] 도 11a에서, 임의의 데이터 차원에서 센싱 값이 D15이고, 윈도우의 크기가 25에서 대체로 학습 결과의 신뢰도가 가장 높은 것으로 가정한다.
- [0233] 본 발명에서는 학습 결과의 신뢰도를 데이터 축약의 적절성 평가를 위한 지표로 사용하므로, 도 11a에서, 센싱 값이 D15일 때 최적의 데이터 축약을 제공하는 윈도우 크기는 25라고 평가할 수 있다. 이때, 최적의 데이터 축약 조건에 대한 평가는 하나의 차원으로 한정되지 않으며, 도 11b에 도시한 바와 같이, 학습 이력이 저장된 모든 데이터 차원에 대해 최적의 데이터 축약을 평가할 수 있다.
- [0234] 하나의 데이터 차원에 대한 최적의 데이터 축약 조건은, 도 11a에서 “지식 증강 구간”으로 예시한 영역에 대해, 도 11c에 예시한 최적화 평가를 통해 최적의 데이터 축약 조건을 찾을 수 있다. 즉, 도 11a에서 “지식 증강 구간”으로 예시한 영역에 포함된 전체 학습 이력을 추출하여, 도 11c에 예시한 것처럼 정렬할 수 있다.
- [0235] 도 11c의 수평축은 도 11a의 수직축과 동일하다. 즉, 도 11c의 수평축은 윈도우 크기를 나타낸다. 도 11c의 수

직측은 RMSE로 나타낸 학습 결과의 신뢰도(또는 정확도)를 의미한다.

- [0236] 도 11a에서 "지식 증강 구간"으로 예시한 영역에 포함된 전체 학습 이력에 대해, RMSE의 크기를 고려하여 2차원 곡선으로 피팅(fitting)을 하면, 데이터 축약을 위한 윈도우의 최적 조건을 평가할 수 있다, 즉, 도 11c에서, 최초로 설정한 축약 기준(50)에서의 윈도우 크기는 20이지만, 학습 이력을 이용하여 피팅한 최적의 축약 기준에서의 최적의 윈도우 크기는 18이 된다.
- [0237] 메타 최적화기(10)는 학습 이력을 이용한 최적의 데이터 축약 조건에 대한 평가를 수행하며, 상기 평가를 이용하여 윈도우 크기를 18로 설정한 새로운 축약 기준 정보를 스키마 정보에 추가한다. 이러한 스키마 정보의 추가 과정에서 사용자의 개입이나 사용자에게 의한 입력이 필요하지 않으므로, 스키마 정보를 자동으로 갱신하는 자가 학습이 실시된다.
- [0238] 본 발명에 의한 자가 학습을 위한 데이터 메타 스케일링 장치 및 방법에서는, 지식 증강 기준을 만족하도록 학습 이력이 충분히 저장된 이후에는, 새로운 학습 이력이 저장될 때마다 도 11a 내지 11c을 참조하여 설명한 지식 증강 프로세스에 따라 지속적인 축약 기준의 최적화를 수행할 수 있게 된다.
- [0239] 이와 같이, 스키마 정보에 포함되어 있는 축약 기준을 갱신하는 과정을 통해 자가 학습을 구성하는 절차를 자동화할 수 있으며, 지속적인 지식 증강 프로세스를 수행하여 데이터 축약을 위한 축약 기준의 최적화를 자동화할 수 있다.
- [0240] 이상에서 설명한 본 발명의 자가 학습을 위한 메타 스케일링 장치 및 방법은, 프로그램으로 구현되어 기록매체에 저장된 후 프로세서에 의해 로딩되어 실행될 수 있다.
- [0241] 본 발명에 따른 기능을 구현하는 프로그램 모듈들, 예를 들면, 메타 최적화기, 축약기, 학습기 및 평가기는 서버 팜(Server Farm)과 같이 네트워크에 걸쳐서 분산형으로 구현될 수 있으며, 혹은 단일의 컴퓨터 장치의 프로세서 내에 임베딩될 수도 있다.
- [0242] 아울러, 본 발명의 자가 학습을 위한 메타 스케일링 장치는, 프로그래머블 프로세서, 컴퓨터 혹은 다중 프로세서나 컴퓨터를 포함하여 데이터를 처리하기 위한 모든 기구, 장치 및 기계에 임베딩될 수 있다.
- [0243] 아울러, 본 발명에 따른 자가 학습을 위한 데이터 메타 스케일링 장치 및 방법은, 예컨대 데이터 서버와 같은 백엔드 컴포넌트를 포함하거나, 어플리케이션 서버와 같은 미들웨어 컴포넌트를 포함하거나, 사용자가 본 명세서에서 설명한 구성과 상호 작용할 수 있는 웹 브라우저나 그래픽 유저 인터페이스를 갖는 클라이언트 컴퓨터와 같은 프론트엔드 컴포넌트 혹은 그러한 백엔드, 미들웨어 혹은 프론트엔드 컴포넌트의 하나 이상의 모든 조합을 더 포함할 수 있다.
- [0245] 이상에서 설명한 실시 예들은 그 일 예로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 발명에 개시된 실시 예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시 예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

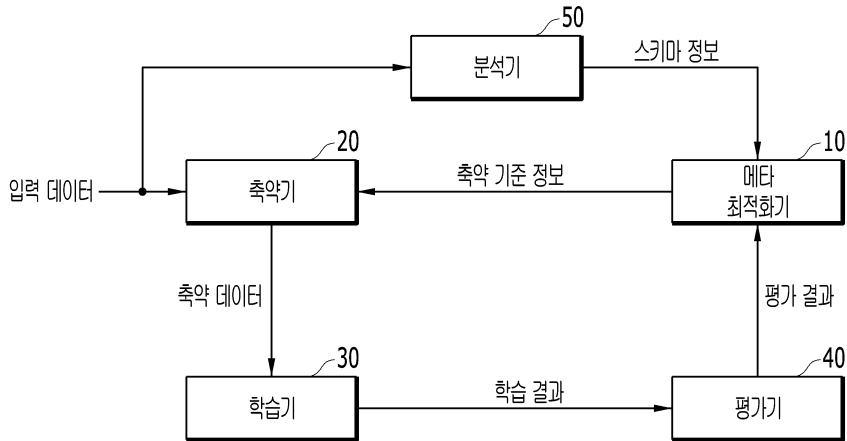
부호의 설명

- [0246] 10, 100: 메타 최적화기
- 110: 복합 차원 데이터 축약기
- 20: 축약기
- 30: 학습기
- 40, 400: 평가기
- 50, 500: 메타 정보 저장소
- 200: 다수의 축약기

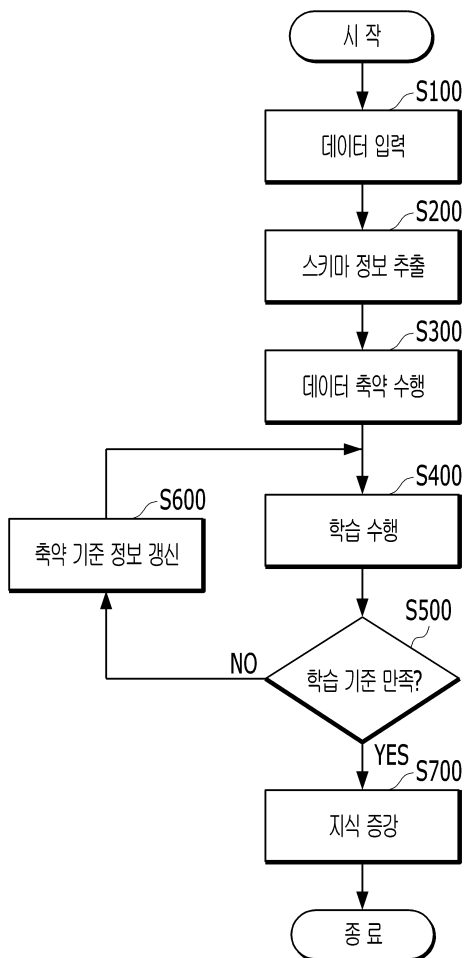
300: 다수의 학습기

도면

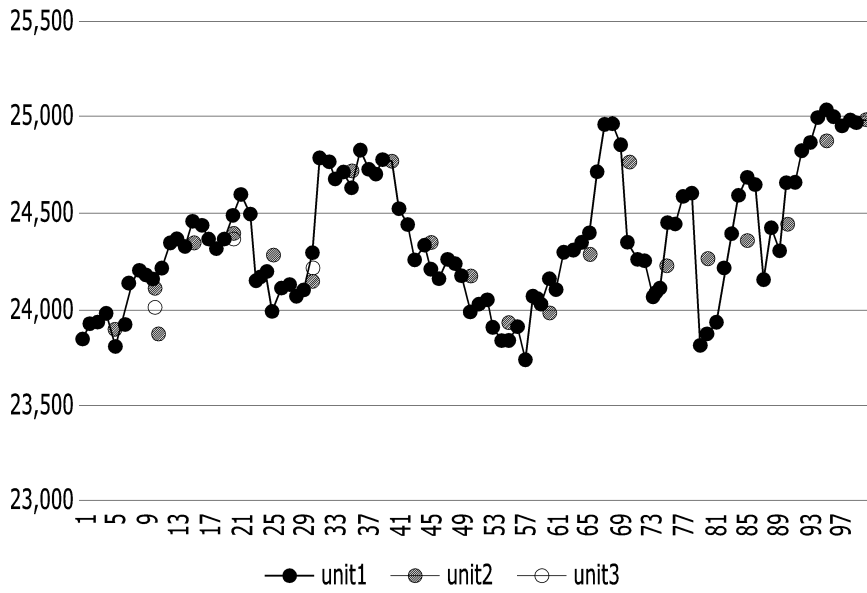
도면1



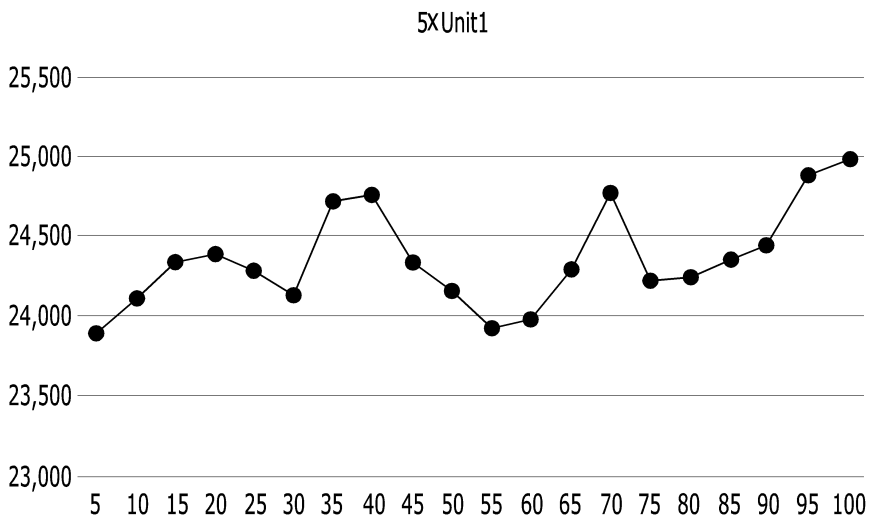
도면2



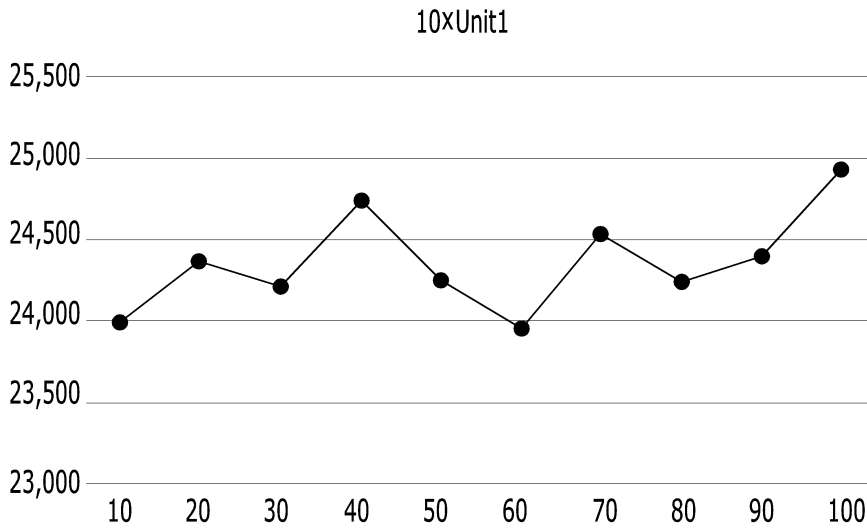
도면3a



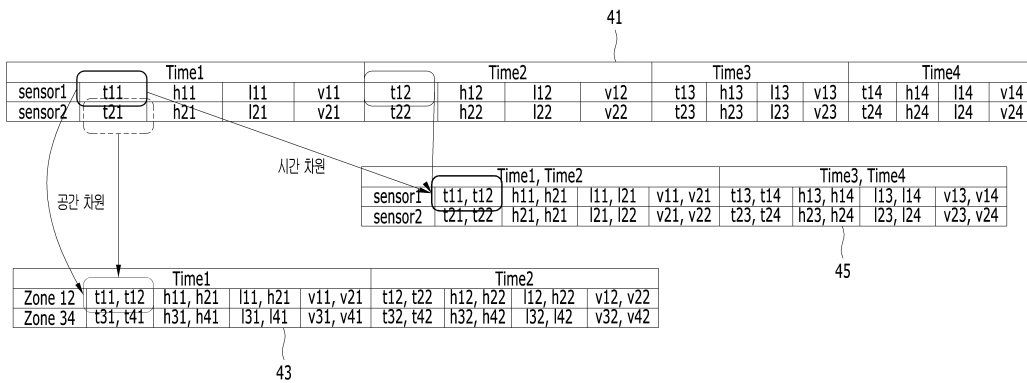
도면3b



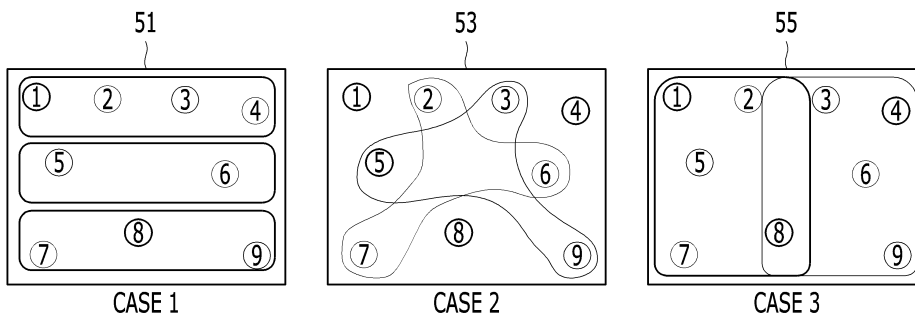
도면3c



도면4



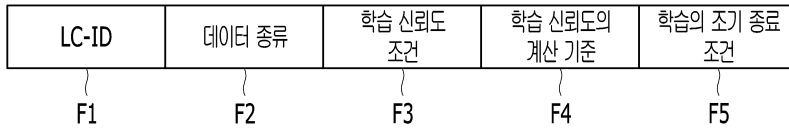
도면5



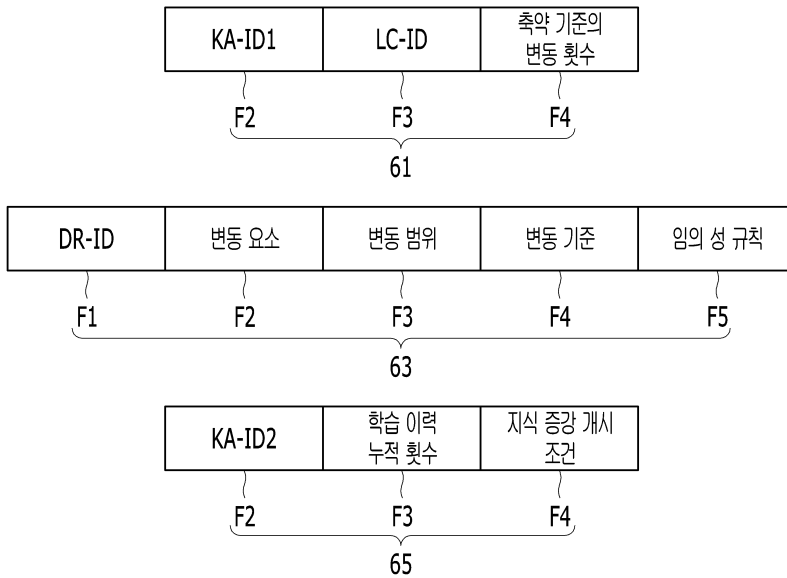
도면6a

DR-ID	데이터 차원	윈도우 종류	윈도우 크기	대표값 선택 기준
F1	F2	F3	F4	F5

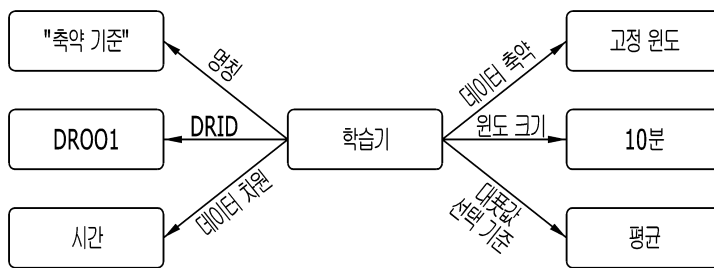
도면6b



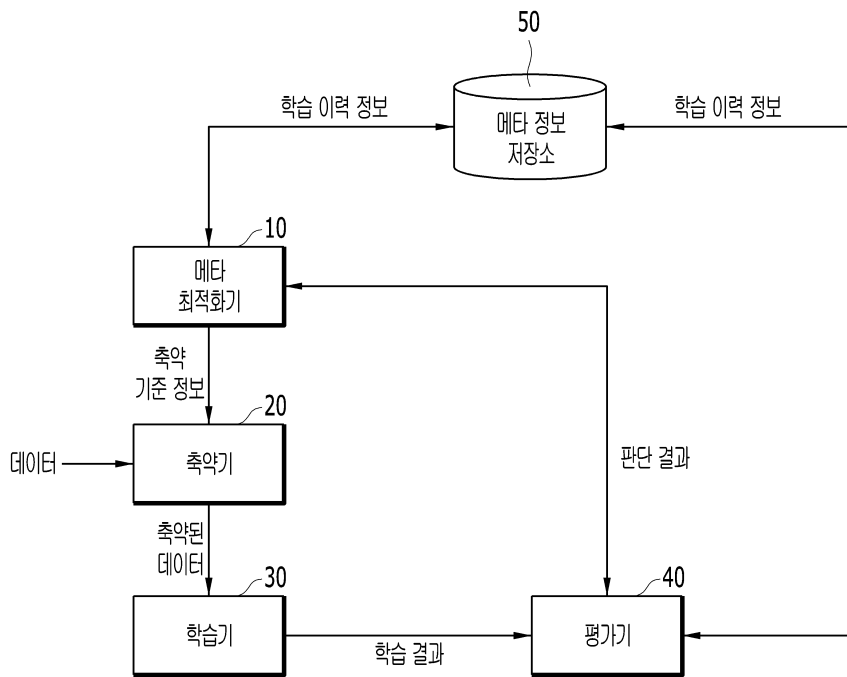
도면6c



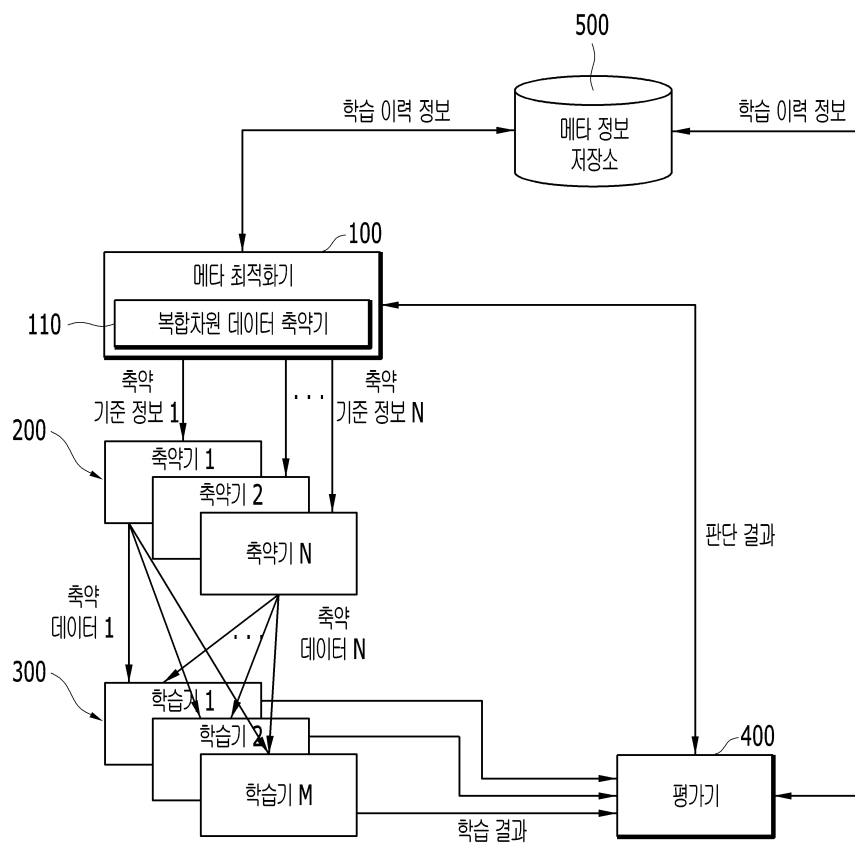
도면7



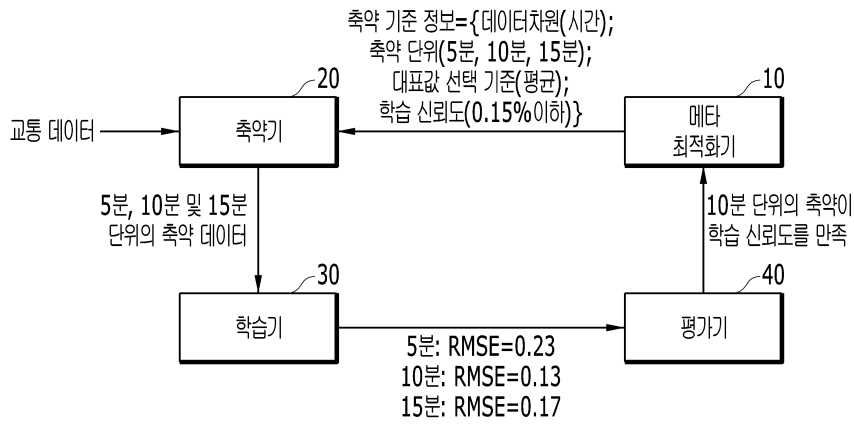
도면8



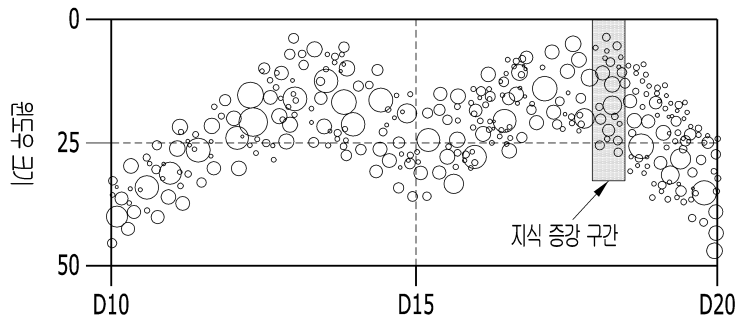
도면9



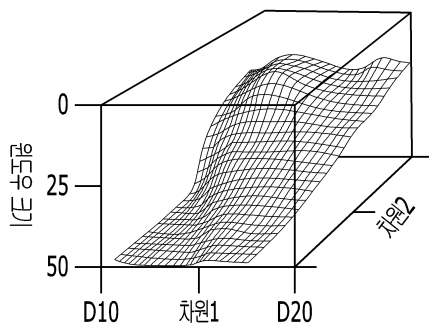
도면10



도면11a



도면11b



도면11c

