(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0074548
(43) 공개일자 2022년06월03일

(51) 국제특허분류(Int. Cl.)
G06F 16/215 (2019.01) *G06F 11/30* (2006.01)
G06F 16/28 (2019.01)
(52) CPC특허분류
G06F 16/215 (2019.01)
G06F 11/3089 (2013.01)
(21) 출원번호 10-2020-0163181
(22) 출원일자 2020년11월27일
심사청구일자 2020년11월27일

(71) 출원인
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
(72) 발명자
반효경
서울특별시 관악구 관악로30길 12, 107동 2001호 (봉천동, 관악우성아파트)
조경운
서울특별시 서초구 방배천로2안길 118(방배동)
(74) 대리인
특허법인 무한

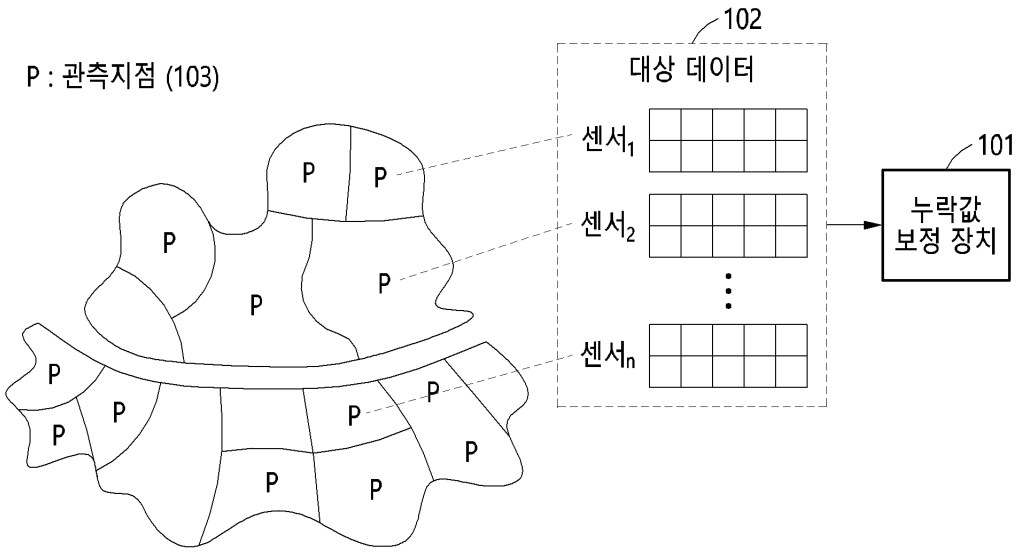
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **센서 데이터의 누락값 보정 방법 및 누락값 보정 장치**

(57) 요 약

본 발명은 테이블로 구성된 센서 데이터에서 누락된 값을 보정하기 위한 누락값 보정 방법 및 누락값 보정 장치에 관한 것으로, 누락값 보정 방법은 센서로부터 수집된 데이터에서 셀 값이 누락된 특정 칼럼을 검출하고, 특정 칼럼을 제외한 유효 칼럼을 이용하여 특정 칼럼의 누락된 셀 값을 보정한다.

대 표 도 - 도1



(52) CPC특허분류
G06F 16/284 (2019.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711117117
과제번호	2020-0-00121-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발(R&D,정보화)
연구과제명	데이터 품질 평가기반 데이터 고도화 및 데이터셋 보정 기술 개발
기 여 율	1/1
과제수행기관명	부산대학교산학협력단
연구기간	2020.04.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

복수의 관측 지점에 설치된 센서로부터 수집한 대상 데이터를 이용하여 테이블로 구성된 검사 데이터를 생성하는 단계;

상기 검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼을 결정하는 단계; 및

상기 검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 상기 특정 칼럼의 누락된 셀 값을 보정하는 단계;

를 포함하는 누락값 보정 방법.

청구항 2

제1항에 있어서,

상기 생성하는 단계는,

상기 복수의 관측 지점에 설치된 센서로부터 시간에 따른 칼럼(Column) 단위의 대상 데이터를 수집하는 단계; 및

상기 수집된 칼럼 단위의 대상 데이터를 순차적으로 취합하여 누락된 셀 값을 보정하기 위한 테이블로 구성된 검사 데이터를 생성하는 단계;

를 포함하는 누락값 보정 방법.

청구항 3

제1항에 있어서,

상기 결정하는 단계는,

상기 테이블을 형성하는 셀들이 나타내는 데이터 타입의 표시 형식을 전처리 하는 단계; 및

상기 표시 형식이 전처리 된 테이블의 셀들 중 셀 값이 누락된 특정 칼럼을 결정하는 단계;

를 포함하는 누락값 보정 방법.

청구항 4

제1항에 있어서,

상기 특정 칼럼의 누락된 셀 값을 보정하는 단계는,

상기 검사 데이터에서 셀 값이 누락된 특정 칼럼의 수집 시간을 추출하는 단계;

시간 유사성에 따른 검사 데이터에서 추출한 수집 시간과 동일 시간에 수집된 적어도 하나의 유효 칼럼을 추출하는 단계; 및

상기 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정하는 단계;

를 포함하는 누락값 보정 방법.

청구항 5

제1항에 있어서,

상기 특정 칼럼의 누락된 셀 값을 보정하는 단계는,

상기 검사 데이터에서 셀 값이 누락된 특정 칼럼을 수집한 센서의 관측 지점을 추출하는 단계;

공간 유사성에 따른 상기 추출한 센서의 관측 지점과 일정 반경 내에 설치된 서로 다른 관측 지점을 판단하는 단계;

상기 검사 데이터에서 서로 다른 관측 지점에 설치된 센서로부터 수집된 적어도 하나의 유효 칼럼을 추출하는 단계; 및

상기 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정하는 단계;

를 포함하는 누락값 보정 방법.

청구항 6

제1항에 있어서,

상기 특정 칼럼의 누락된 셀 값을 보정하는 단계는,

상기 유효 칼럼들에 따른 유효 행에 대한 데이터 패턴에 기초하여 특정 칼럼의 누락된 셀 값을 보정하는 누락값 보정 방법.

청구항 7

프로세서를 포함하는 누락값 보정 장치에 있어서,

상기 프로세서는,

복수의 관측 지점에 설치된 센서로부터 수집한 대상 데이터를 이용하여 테이블로 구성된 검사 데이터를 생성하고,

상기 검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼을 결정하고,

상기 검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 상기 특정 칼럼의 누락된 셀 값을 보정하는 누락값 보정 장치.

청구항 8

제7항에 있어서,

상기 프로세서는,

상기 복수의 관측 지점에 설치된 센서로부터 시간에 따른 칼럼 단위의 대상 데이터를 수집하고,

상기 수집된 칼럼 단위의 대상 데이터를 순차적으로 취합하여 누락된 셀 값을 보정하기 위한 테이블로 구성된 검사 데이터를 생성하는 누락값 보정 장치.

청구항 9

제7항에 있어서,

상기 프로세서는,

상기 테이블을 형성하는 셀들이 나타내는 데이터 타입의 표시 형식을 전처리를 수행하고,

상기 표시 형식이 전처리 된 테이블의 셀들 중 셀 값이 누락된 특정 칼럼을 결정하는 누락값 보정 장치.

청구항 10

제7항에 있어서,

상기 프로세서는,

상기 검사 데이터에서 셀 값이 누락된 특정 칼럼의 수집 시간을 추출하고,

시간 유사성에 따른 검사 데이터에서 추출한 수집 시간과 동일 시간에 수집된 적어도 하나의 유효 칼럼을 추출

하고,

상기 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정하는 누락값 보정 장치.

청구항 11

제1항에 있어서,

상기 프로세서는,

상기 검사 데이터에서 셀 값이 누락된 특정 칼럼을 수집한 센서의 관측 지점을 추출하고,

공간 유사성에 따른 상기 추출한 센서의 관측 지점과 일정 반경 내에 설치된 서로 다른 관측 지점을 판단하고,

상기 검사 데이터에서 서로 다른 관측 지점에 설치된 센서로부터 수집된 적어도 하나의 유효 칼럼을 추출하고,

상기 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정하는 누락값 보정 장치.

청구항 12

제7항에 있어서,

상기 프로세서는,

상기 유효 칼럼들에 따른 유효 행에 대한 데이터 패턴에 기초하여 특정 칼럼의 누락된 셀 값을 보정하는 누락값 보정 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 센서 데이터의 누락값 보정 방법 및 누락값 보정 장치에 관한 것으로, 보다 구체적으로는 센서로부터 수집된 센서 데이터에 포함된 누락 값을 자동으로 보정하여 고품질의 데이터를 도출하기 위한 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 일반적으로, 테이블 데이터는 열과 행으로 구성된 데이터로써, 관계형 데이터베이스에서 주로 쓰는 표현 형식이며, 다수의 속성을 가지는 많은 개수의 항목들을 표시하는 방법이다. 테이블 데이터의 각 행은 데이터 항목이며, 열을 각 항목이 가지는 속성으로서 테이블의 각 셀은 항목의 속성값으로 볼 수 있다.

[0003] 테이블의 각 셀 중 일부 데이터가 누락된 테이블 데이터의 경우, 데이터 분석을 하거나 또는, 기계 학습을 위한 훈련 데이터를 사용할 때, 누락된 데이터를 처리하기 위하여 예외 처리를 해야 하는 부담이 발생한다. 이때, 1개 셀의 누락으로 인하여 전체 행 데이터를 사용하지 못하는 경우도 발생할 수 있다.

[0004] 예외 처리를 제공하는 경우, 해당 데이터 항목을 모두 배제하거나 평균값 혹은 단순 유추값 등으로 대치되는데, 데이터 항목을 모두 배제하는 경우에는 유효한 다른 칼럼의 값들을 데이터 분석에서 제외하게 되는 문제가 발생한다. 또한, 평균값 혹은 단순 유추값으로 대치되는 경우에는 누락된 데이터에 대해 특정값으로 평향되는 문제가 발생한다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 테이블로 구성된 데이터에서 셀 값이 누락된 특정 칼럼을 검출하고, 특정 칼럼을 제외한 유효 칼럼에 따른 유효 행을 이용하여 누락된 셀 값을 보정함으로써, 특정 칼럼의 오류로 인한 누락된 값이 포함된 전체 행이 데이터 분석에서 제외하는 것을 방지하는 장치 및 방법을 제공한다.

[0006] 본 발명은 누락된 셀 값을 포함하는 특정 칼럼을 제외한 유효 칼럼의 모든 행을 학습한 후, 학습 결과에 기반한 특정 칼럼의 누락된 셀 값을 보정함으로써, 테이블을 구성하는 행 간의 연관성에 따른 특정 값으로 편향되는 것

을 방지하는 장치 및 방법을 제공한다.

[0007] 본 발명은 특정 칼럼의 누락된 셀 값을 보정함에 있어, 데이터를 수집하는 과정에서의 공간 유사성 및 시간 유사성을 고려함으로써, 시간 및 장소의 특성에 따른 누락된 셀 값을 보정하는 장치 및 방법을 제공한다.

과제의 해결 수단

[0008] 본 발명의 일실시예에 따른 누락값 보정 방법은 복수의 관측 지점에 설치된 센서로부터 수집한 대상 데이터를 이용하여 테이블로 구성된 검사 데이터를 생성하는 단계; 상기 검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼을 결정하는 단계; 및 상기 검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 상기 특정 칼럼의 누락된 셀 값을 보정하는 단계;를 포함할 수 있다.

[0009] 본 발명의 실시예에 따른 생성하는 단계는, 상기 복수의 관측 지점에 설치된 센서로부터 시간에 따른 칼럼(Column) 단위의 대상 데이터를 수집하는 단계; 및 상기 수집된 칼럼 단위의 대상 데이터를 순차적으로 취합하여 누락된 셀 값을 보정하기 위한 테이블로 구성된 검사 데이터를 생성하는 단계;를 포함할 수 있다.

[0010] 본 발명의 실시예에 따른 결정하는 단계는, 상기 테이블을 형성하는 셀들이 나타내는 데이터 타입의 표시 형식을 전처리 하는 단계; 및 상기 표시 형식이 전처리 된 테이블의 셀들 중 셀 값이 누락된 특정 칼럼을 결정하는 단계;를 포함할 수 있다.

[0011] 본 발명의 실시예에 따른 특정 칼럼의 누락된 셀 값을 보정하는 단계는, 상기 검사 데이터에서 셀 값이 누락된 특정 칼럼의 수집 시간을 추출하는 단계; 시간 유사성에 따른 검사 데이터에서 추출한 수집 시간과 동일 시간에 수집된 적어도 하나의 유효 칼럼을 추출하는 단계; 및 상기 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정하는 단계;를 포함할 수 있다.

[0012] 본 발명의 실시예에 따른 특정 칼럼의 누락된 셀 값을 보정하는 단계는, 상기 검사 데이터에서 셀 값이 누락된 특정 칼럼을 수집한 센서의 관측 지점을 추출하는 단계; 공간 유사성에 따른 상기 추출한 센서의 관측 지점과 일정 반경 내에 설치된 서로 다른 관측 지점을 판단하는 단계; 상기 검사 데이터에서 서로 다른 관측 지점에 설치된 센서로부터 수집된 적어도 하나의 유효 칼럼을 추출하는 단계; 및 상기 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정하는 단계;를 포함할 수 있다.

[0013] 본 발명의 실시예에 따른 특정 칼럼의 누락된 셀 값을 보정하는 단계는, 상기 유효 칼럼들에 따른 유효 행에 대한 데이터 패턴에 기초하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0014] 본 발명의 일실시예에 따른 프로세서를 포함하는 누락값 보정 장치에 있어서, 상기 프로세서는, 복수의 관측 지점에 설치된 센서로부터 수집한 대상 데이터를 이용하여 테이블로 구성된 검사 데이터를 생성하고, 상기 검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼을 결정하고, 상기 검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 상기 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0015] 본 발명의 실시예에 따른 프로세서는, 상기 복수의 관측 지점에 설치된 센서로부터 시간에 따른 칼럼 단위의 대상 데이터를 수집하고, 상기 수집된 칼럼 단위의 대상 데이터를 순차적으로 취합하여 누락된 셀 값을 보정하기 위한 테이블로 구성된 검사 데이터를 생성할 수 있다.

[0016] 본 발명의 일실시예에 따른 프로세서는, 상기 테이블을 형성하는 셀들이 나타내는 데이터 타입의 표시 형식을 전처리를 수행하고, 상기 표시 형식이 전처리 된 테이블의 셀들 중 셀 값이 누락된 특정 칼럼을 결정할 수 있다.

[0017] 본 발명의 일실시예에 따른 프로세서는, 상기 검사 데이터에서 셀 값이 누락된 특정 칼럼의 수집 시간을 추출하고, 시간 유사성에 따른 검사 데이터에서 추출한 수집 시간과 동일 시간에 수집된 적어도 하나의 유효 칼럼을 추출하고, 상기 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0018] 본 발명의 일실시예에 따른 프로세서는, 상기 검사 데이터에서 셀 값이 누락된 특정 칼럼을 수집한 센서의 관측 지점을 추출하고, 공간 유사성에 따른 상기 추출한 센서의 관측 지점과 일정 반경 내에 설치된 서로 다른 관측 지점을 판단하고, 상기 검사 데이터에서 서로 다른 관측 지점에 설치된 센서로부터 수집된 적어도 하나의 유효 칼럼을 추출하고, 상기 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0019] 본 발명의 일실시예에 따른 프로세서는, 상기 유효 칼럼들에 따른 유효 행에 대한 데이터 패턴에 기초하여 특정

칼럼의 누락된 셀 값을 보정할 수 있다.

발명의 효과

[0020] 본 발명의 일실시예에 따른 누락값 보정 방법은 테이블로 구성된 수집 데이터에서 셀 값이 누락된 특정 칼럼을 검출하고, 특정 칼럼을 제외한 유효 칼럼에 따른 유효 행을 이용하여 누락된 셀 값을 보정함으로써, 특정 칼럼의 오류로 인한 누락된 값이 포함된 전체 행이 데이터 분석에서 제외하는 것을 방지할 수 있다.

[0021] 본 발명의 일실시예에 따른 누락값 보정 방법은 누락된 셀 값을 포함하는 특정 칼럼을 제외한 유효 칼럼의 모든 행을 학습한 후, 학습 결과에 기반한 특정 칼럼의 누락된 셀 값을 보정함으로써, 테이블을 구성하는 행 간의 연관성에 따른 특정 값으로 편향되는 것을 방할 수 있다.

[0022] 본 발명의 일실시예에 따른 누락값 보정 방법은 특정 칼럼의 누락된 셀 값을 보정함에 있어, 데이터를 수집하는 과정에서의 공간 유사성 및 시간 유사성을 고려함으로써, 시간 및 장소의 특성에 따른 누락된 셀 값을 보정할 수 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 일실시예에 따른 테이블로 구성된 데이터의 누락값을 보정하기 위한 전체적인 동작을 설명하기 위해 도시한 도면이다.

도 2는 본 발명의 일실시예에 따른 누락값 보정 장치의 세부적인 동작을 설명하기 위해 도시한 도면이다.

도 3은 본 발명의 일실시예에 따른 데이터의 누락값을 예측하기 위한 과정을 설명하기 위해 도시한 도면이다.

도 4는 본 발명의 일실시예에 따른 데이터의 누락값을 보정하는 과정을 설명하기 위해 도시한 도면이다.

도 5는 본 발명의 일실시예에 따른 누락값 보정 방법을 설명하기 위해 도시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0024] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 구현될 수 있다. 따라서, 실제 구현되는 형태는 개시된 특정 실시예로만 한정되는 것이 아니며, 본 명세서의 범위는 실시예들로 설명한 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.

[0025] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.

[0026] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.

[0027] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0028] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0029] 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일 구성 요소는 동일 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.

[0031] 도 1은 본 발명의 일실시예에 따른 테이블로 구성된 센서 데이터의 누락값을 보정하기 위한 전체적인 동작을 설명하기 위해 도시한 도면이다.

[0032] 도 1을 참고하면, 누락값 보정 장치(101)는 테이블로 구성된 센서 데이터에서 누락된 셀 값을 자동으로 보정할 수 있다. 자세하게, 누락값 보정 장치(101)는 셀 값이 누락된 특정 칼럼의 유효한 값을 보유한 모든 행들에 대한 센서 데이터의 패턴을 학습하여 누락된 행의 유사한 값을 추출할 수 있다. 여기서, 셀 값이 누락된 특정 칼럼의 유효한 값은 셀 값이 누락된 특정 칼럼을 제외한 나머지 유효 칼럼에서 누락된 셀과 동일한 행의 유효한 값을 의미할 수 있다. 누락값 보정 장치(101)는 행의 유효한 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0033] 이를 위해, 누락값 보정 장치(101)는 복수의 관측 지점(103)에 설치된 센서로부터 수집한 대상 데이터(102)를 이용하여 테이블로 구성된 검사 데이터를 생성할 수 있다. 누락값 보정 장치(101)는 테이블을 형성하는 셀들이 나타내는 데이터 타입의 표시 형식을 전처리 과정을 수행할 수 있다. 여기서, 전처리 과정은 각 칼럼의 자료형이 수치값이 아닌 경우, 이를 수치값으로 변환하는 과정을 의미할 수 있다. 일례로, 각 칼럼의 자료형이 범주형 데이터의 경우, 가능한 범주 문자열을 순차적인 숫자값으로 변환할 수 있다. 또한, 각 칼럼의 자료형이 텍스트의 경우, 용어 리스트에 기재된 개별 식별자 번호를 기반으로 해당 텍스트를 수치값으로 변환할 수 있다. 또한, 모든 수치값은 정규화를 하여 '0'에서 '1'사이의 값으로 변환할 수 있다. 이때, 수치화가 불가능한 칼럼은 제외한다.

[0034] 누락값 보정 장치(101)는 검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼을 결정할 수 있다. 누락값 보정 장치(101)는 검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 상기 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0036] 도 2는 본 발명의 일실시예에 따른 누락값 보정 장치의 세부적인 동작을 설명하기 위해 도시한 도면이다.

[0037] 도 2를 참고하면, 누락값 보정 장치(201)는 누락값 보정을 위한 프로세서(202)를 포함할 수 있다. 프로세서(202)는 센서로부터 수집된 데이터에서 셀 값이 누락된 특정 칼럼을 검출하고, 특정 칼럼을 제외한 유효 칼럼을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다. 이를 위해, 프로세서(202)는 복수의 관측 지점에 설치된 센서(204)로부터 수집한 대상 데이터(205)를 수신할 수 있다. 프로세서(202)는 수신한 대상 테이블(205)를 취합하여 하나의 테이블로 구성된 검사 데이터(203)를 생성할 수 있다.

[0038] 프로세서(202)는 센서로부터 시간에 따른 칼럼(Column) 단위의 대상 데이터(205)를 수집할 수 있다. 프로세서(202)는 수집된 칼럼 단위의 대상 데이터(205)를 순차적으로 취합하여 누락된 셀 값을 보정하기 위한 테이블로 구성된 검사 데이터(203)를 생성할 수 있다. 프로세서(202)는 센서로부터 대상 데이터(205)를 수집한 시간을 기준으로 순서에 따라 차례대로 취합하여 테이블로 구성된 검사 데이터(203)를 생성할 수 있다.

[0039] 프로세서(202)는 검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼을 결정할 수 있다. 프로세서(202)는 테이블을 형성하는 셀들이 나타내는 데이터 타입의 표시 형식을 전처리 과정을 수행할 수 있다. 프로세서(202)는 표시 형식이 전처리 된 테이블의 셀들 중 셀 값이 누락된 특정 칼럼을 결정할 수 있다.

[0040] 프로세서(202)는 검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0042] 도 3은 본 발명의 일실시예에 따른 데이터의 누락값을 예측하기 위한 과정을 설명하기 위해 도시한 도면이다.

[0043] 도 3의 (a)를 살펴보면, 누락값 보정 장치는 검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값(302)이 누락된 특정 칼럼(301)을 결정할 수 있다.

[0044] 자세하게, 누락값 보정 장치는 테이블을 형성하는 셀들이 나타내는 데이터 타입의 표시 형식을 전처리 과정을 수행할 수 있다. 여기서, 전처리 과정은 테이블로 구성된 검사 데이터의 각 칼럼이 나타내는 데이터 타입을 보정하는 과정일 수 있다. 누락값 보정 장치는 각 칼럼의 데이터 타입을 식별하고, 식별된 데이터 타입을 수치값으로 변환하는 작업을 수행할 수 있다. 일례로, 누락값 보정 장치는 칼럼의 데이터 타입이 스트링 타입(string type)인 경우, 스트링 타입으로 표현된 문자열을 순차적인 숫자값 또는, 개별 식별자 번호에 기초한 수치값으로 변환하는 작업을 수행할 수 있다.

[0045] 누락값 보정 장치는 표시 형식이 전처리 된 테이블의 셀들 중 셀 값(302)이 누락된 특정 칼럼(301)을 결정할 수 있다. 누락값 보정 장치는 테이블의 셀들 중 공백값 또는, 문자열의 수치값이 표현된 셀을 검출하고, 해당 셀

에 대하여 값이 누락된 것으로 판단할 수 있다.

[0046] 도 3의 (b)를 살펴보면, 누락값 보정 장치는 검사 데이터에서 특정 칼럼(301)을 제외한 유효 칼럼들(303)에 따른 유효 행(304)을 이용하여 특정 칼럼(301)의 누락된 셀 값(302)을 보정할 수 있다. 누락값 보정 장치는 누락값 보정 장치는 순 전파 망(FFN) / 역 전파 망(BPN) 또는 회귀 분석을 통한 학습 모델을 활용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0047] 누락값 보정 장치는 유효 칼럼들(303)에 따른 유효 행(304)의 각 셀 값에 대하여 순 전파 망(FFN) / 역 전파 망(BPN) 또는 회귀 분석을 통한 학습 모델에 적용할 수 있다. 누락값 보정 장치는 이에 따른 결과에 기초하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0048] 이때, 누락값 보정 장치는 특정 칼럼의 누락된 셀 값을 보정함에 있어, 데이터를 수집하는 과정에서의 공간 유사성 및 시간 유사성을 고려함으로써, 시간 및 장소의 특성에 따른 누락된 셀 값을 보정할 수 있다.

[0049] ① 시간 유사성

[0050] 누락값 보정 장치는 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼이 결정되면, 검사 데이터에서 셀 값이 누락된 특정 칼럼의 수집 시간을 추출할 수 있다. 이때, 수집 시간은 관측 지점에 설치된 센서로부터 일정 주기에 따라 데이터가 수집된 시간일 수 있다. 누락값 보정 장치는 시간 유사성에 따른 검사 데이터에서 추출한 수집 시간과 동일 시간에 수집된 적어도 하나의 유효 칼럼을 추출할 수 있다. 여기서, 추출된 유효 칼럼은 이전에 수집된 데이터로써, 누락된 셀 값이 존재하지 않으며, 누락된 셀 값을 보정하기 위해 활용되는 유효한 값을 갖는 칼럼을 의미할 수 있다. 일례로, 누락값 보정 장치는 현재를 기준으로 일간, 주간, 월간, 년간을 기준으로 동일 시간에 수집된 적어도 하나의 유효 칼럼을 추출할 수 있다.

[0051] 그리고, 누락값 보정 장치는 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다. 누락값 보정 장치는 적어도 하나의 유효 칼럼에 포함된 셀 값들 간에 차이 값, 중앙 값, 평균 값을 결정할 수 있다. 누락값 보정 장치는 결정된 각각의 값 들로부터 최적인 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0052] ② 공간 유사성

[0053] 누락값 보정 장치는 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼이 결정되면, 검사 데이터에서 셀 값이 누락된 특정 칼럼을 수집한 센서가 설치된 관측 지점을 판단할 수 있다. 누락값 보정 장치는 공간 유사성에 따른 추출한 센서의 관측 지점과 일정 반경 내에 설치된 서로 다른 관측 지점을 판단할 수 있다. 누락값 보정 장치는 검사 데이터에서 서로 다른 관측 지점에 설치된 센서로부터 수집된 적어도 하나의 유효 칼럼을 추출할 수 있다.

[0054] 일례로, 셀 값이 누락된 특정 칼럼을 수집한 관측 지점이 강남이며, 인접한 관측 지점이 서초, 은평일 때, 누락값 보정 장치는 검사 데이터에서 서초, 은평에서 수집된 데이터에 해당하는 유효 칼럼을 추출할 수 있다.

[0056] 도 4는 본 발명의 일실시예에 따른 데이터의 누락값을 보정하는 과정을 설명하기 위해 도시한 도면이다.

[0057] 도 4를 참고하면, 누락값 보정 장치는 검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다. 이를 위해, 누락값 보정 장치는 누락값 보정 장치는 순 전파 망(FFN: Feed Forward Network) 및 역 전파 망(BPN: Back Propagation Network)을 활용할 수 있다. 또한, 누락값 보정 장치는 회귀 분석을 통한 학습 모델을 활용할 수 있다.

[0058] 누락값 보정 장치는 순 전파 망 및 역 전파 망에 따른 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행에 대한 데이터 패턴을 결정할 수 있다. 누락값 보정 장치는 데이터 패턴에 기초하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0059] ① 순 전파 망 및 역 전파 망을 활용하는 방법

[0060] 누락값 보정 장치는 누락된 셀의 개수에 따른 순 전파 망 및 역 전파 망을 활용하여 누락된 셀을 보유한 행들에 대한 누락된 셀 값을 보정할 수 있다. 여기서, 순 전파 망의 입력 개수는 특정 칼럼을 제외한 유효 칼럼의 개수이며, 역 전파 망의 출력 개수는 누락된 셀의 개수일 수 있다. 본 발명은 검사 데이터를 순 전파 망 및 역 전파 망을 학습하기 위한 훈련 데이터로 이용할 수 있으며, 훈련 데이터는 유효 칼럼을 보유한 모든 행의 개수

를 포함할 수 있다.

[0061] 순 전파 망의 입력은 검사 데이터에서 누락된 셀을 포함하는 행의 유효 칼럼에 대한 속성값이며, 역 전파 망의 출력은 특정 칼럼의 누락된 셀 값으로 예측된 값일 수 있다.

[0062] ② 회귀 분석을 통한 학습 모델을 활용하는 방법

[0063] 누락값 보정 장치는 독립변수(independent variable)가 종속변수(dependent variable)에 미치는 영향을 확인하고자 사용하는 회귀 분석을 이용하여 특정 칼럼의 누락된 셀 값을 예측 및 보정할 수 있다.

[0064] 누락값 보정 장치는 값이 누락된 셀이 나타내는 항목을 판단하고, 검사 데이터에서 판단된 항목을 제외한 나머지 항목을 구분할 수 있다. 누락값 보정 장치는 나머지 항목 및 각 항목의 행을 독립 변수로 설정하고, 누락된 셀의 행을 종속 변수로 설정할 수 있다. 누락값 보정 장치는 독립 변수와 종속 변수에 따른 회귀 분석을 통한 학습 모델을 생성할 수 있다. 누락값 보정 장치는 학습 모델을 통한 이용하여 특정 칼럼의 누락된 셀 값을 예측 및 보정할 수 있다.

[0066] 도 5는 본 발명의 일실시예에 따른 누락값 보정 방법을 설명하기 위해 도시한 흐름도이다.

[0067] 단계(501)에서 누락값 보정 장치는 복수의 관측 지점에 설치된 센서로부터 수집한 대상 데이터를 이용하여 테이블로 구성된 검사 데이터를 생성할 수 있다. 여기서, 누락값 보정 장치는 복수의 관측 지점에 설치된 센서로부터 시간에 따른 칼럼(Column) 단위의 대상 데이터를 수집할 수 있다. 누락값 보정 장치는 수집된 칼럼 단위의 대상 데이터를 순차적으로 취합하여 누락된 셀 값을 보정하기 위한 테이블로 구성된 검사 데이터를 생성할 수 있다.

[0068] 단계(502)에서 누락값 보정 장치는 검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼을 결정할 수 있다. 누락값 보정 장치는 테이블을 형성하는 셀들이 나타내는 데이터 타입의 표시 형식을 전처리 과정을 수행할 수 있다. 누락값 보정 장치는 표시 형식이 전처리 된 테이블의 셀들 중 셀 값이 누락된 특정 칼럼을 결정할 수 있다.

[0069] 단계(503)에서 누락값 보정 장치는 검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다. 누락값 보정 장치는 검사 데이터를 생성하기 위해 센서로부터 수집한 대상 데이터의 시간 유사성 또는, 대상 데이터를 수집한 센서가 설치된 위치의 공간 유사성을 고려할 수 있다.

[0070] 누락값 보정 장치는 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값을 이용하여 특정 칼럼의 누락된 셀 값을 보정할 수 있다. 누락값 보정 장치는 추출된 적어도 하나의 유효 칼럼에 따른 유효 행의 셀 값에 대한 회귀 모델을 통한 학습을 수행하고, 학습 결과에 따른 특정 칼럼의 누락된 셀 값을 보정할 수 있다.

[0072] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0073] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나

처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0074] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있으며 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0075] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0076] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0077] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

부호의 설명

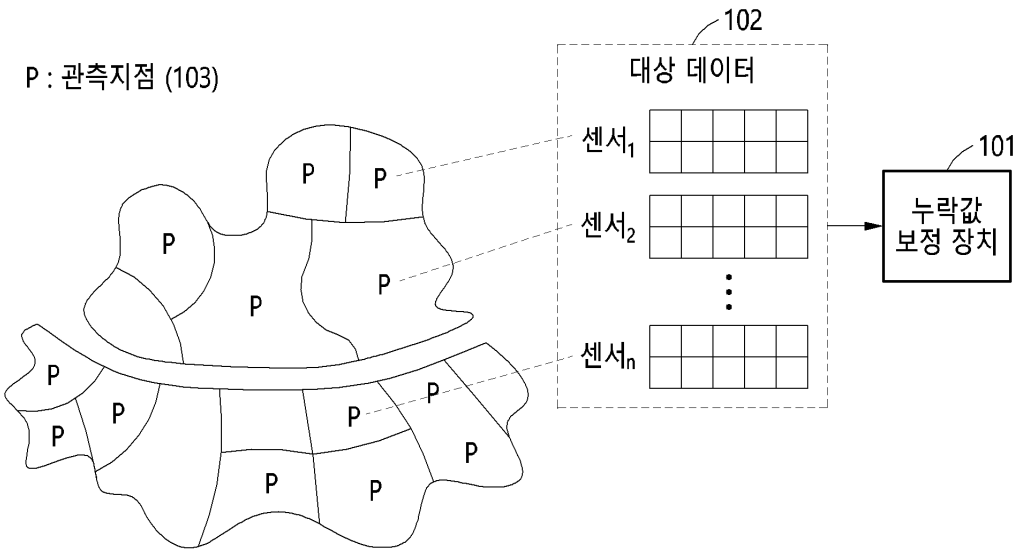
[0078] 101: 누락값 보정 장치

102: 대상 데이터

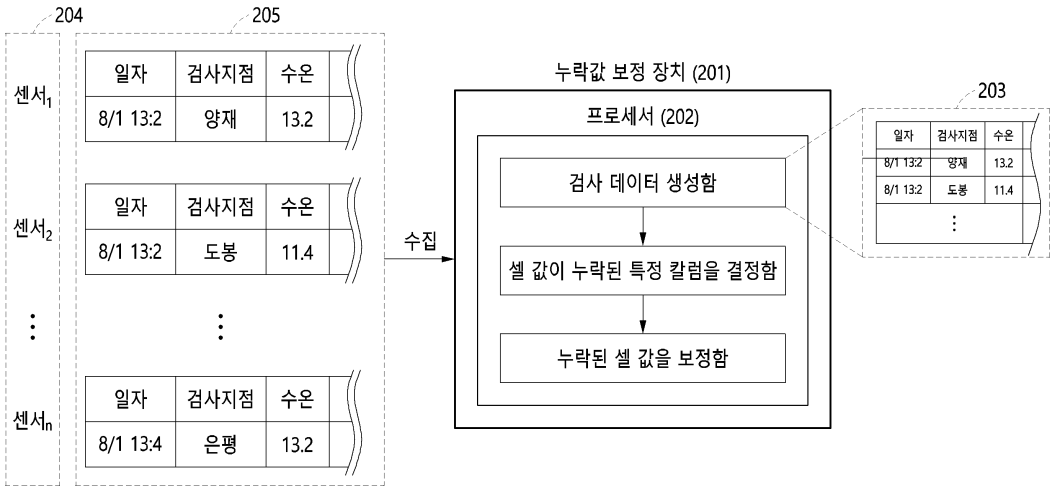
103: 관측 지점

도면

도면1



도면2


204
205
센서1
센서2
센서n
일자
검사지점
수온
8/1 13:2
양재
13.2
8/1 13:2
도봉
11.4
8/1 13:4
은평
13.2
수집
누락값 보정 장치 (201)
프로세서 (202)
검사 데이터 생성함
셀 값이 누락된 특정 칼럼을 결정함
누락된 셀 값을 보정함
203

도면3

일자	검사지점	수온	PH	전기전도도	연소농도	산소농도	기온
08/01 13:12	양재	13.2	6.3	0.9	1.19	12.1	20.1
08/01 13:12	도봉	11.4	6.9	0.9	1.24	12.2	20.1
08/01 13:12	서초	14.2	6.2	0.8	1.23	12.3	20.2
08/01 13:12	은평	13.2	6.8	1.1	1.09	12.2	21.1
08/01 13:12	종로	11.9	7.0	0.9	-	12.1	20.1
08/01 13:12	서초	12.4	7.1	1.1	1.20	12.2	20.4
08/01 13:12	강남	13.9	6.8	1.2	1.22	12.1	20.2

302 특정 칼럼 (301)

(a)

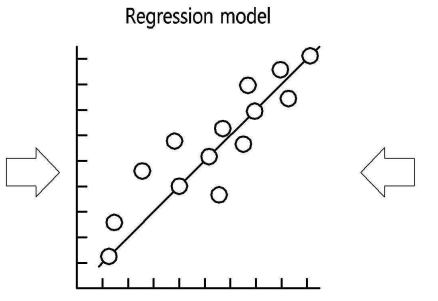
일자	검사지점	수온	PH	전기전도도	연소농도	산소농도	기온
08/01 13:12	양재	13.2	6.3	0.9	1.19	12.1	20.1
08/01 13:12	도봉	11.4	6.9	0.9	1.24	12.2	20.1
08/01 13:12	서초	14.2	6.2	0.8	1.23	12.3	20.2
08/01 13:12	은평	13.2	6.8	1.1	1.09	12.2	21.1
08/01 13:12	종로	11.9	7.0	0.9	-	12.1	20.1
08/01 13:12	서초	12.4	7.1	1.1	1.20	12.2	20.4
08/01 13:12	강남	13.9	6.8	1.2	1.22	12.1	20.2

유효 칼럼 (303) 유효행 (304)

(b)

도면4

수온	PH	전기전도도	산소농도	기온
13.2	6.3	0.9	12.1	20.1
11.4	6.9	0.9	12.2	20.1
14.2	6.2	0.8	12.3	20.2
13.2	6.8	1.1	12.2	21.1
11.9	7.0	0.9	12.1	20.1
12.4	7.1	1.1	12.2	20.4
13.9	6.8	1.2	12.1	20.2



종속변수

염소농도
1.19
1.24
1.23
1.09
1.20
1.22

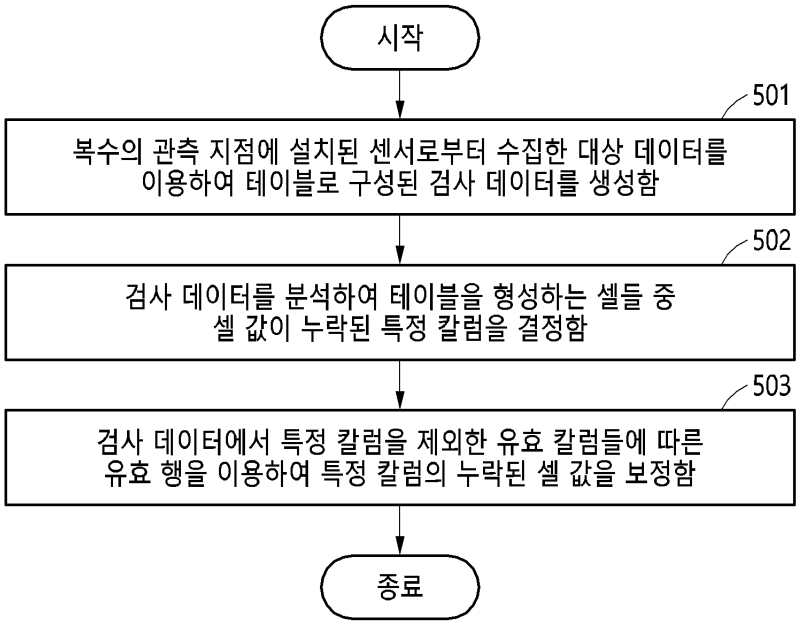
Fitting

일자	검사지점	수온	PH	전기전도도	연소농도	산소농도	기온
08/01 13:13	종로	11.9	7.0	0.9	-	12.1	20.1

결측치 예측 엔진

염소농도
1.19

도면5


시작
501
복수의 관측 지점에 설치된 센서로부터 수집한 대상 데이터를 이용하여 테이블로 구성된 검사 데이터를 생성함
502
검사 데이터를 분석하여 테이블을 형성하는 셀들 중 셀 값이 누락된 특정 칼럼을 결정함
503
검사 데이터에서 특정 칼럼을 제외한 유효 칼럼들에 따른 유효 행을 이용하여 특정 칼럼의 누락된 셀 값을 보정함
종료