



등록특허 10-2345302



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년12월31일
(11) 등록번호 10-2345302
(24) 등록일자 2021년12월27일

- (51) 국제특허분류(Int. Cl.)
G06F 16/21 (2019.01) *G06F 16/22* (2019.01)
G06F 16/28 (2019.01) *G06Q 10/10* (2012.01)
- (52) CPC특허분류
G06F 16/217 (2019.01)
G06F 16/221 (2019.01)
- (21) 출원번호 10-2020-0023603
- (22) 출원일자 2020년02월26일
심사청구일자 2020년02월26일
- (65) 공개번호 10-2021-0027024
- (43) 공개일자 2021년03월10일
- (30) 우선권주장
JP-P-2019-159980 2019년09월02일 일본(JP)
- (56) 선행기술조사문헌
KR1020190029671 A*
(뒷면에 계속)

전체 청구항 수 : 총 9 항

- (73) 특허권자
가부시키가이샤 히타치세이사쿠쇼
일본국 도쿄도 지요다구 마루노우치 1초메 6반 6
고
- (72) 발명자
야마모토, 히데노리
일본 1008280 도쿄도 지요다구 마루노우찌 1초메
6-6 가부시키가이샤 히타치 세이사쿠쇼 내
쓰노, 다까시
- 일본 1008280 도쿄도 지요다구 마루노우찌 1초메
6-6 가부시키가이샤 히타치 세이사쿠쇼 내
사이또, 모또노부
- 일본 1008280 도쿄도 지요다구 마루노우찌 1초메
6-6 가부시키가이샤 히타치 세이사쿠쇼 내

- (74) 대리인
장수길, 박상돈, 이중희

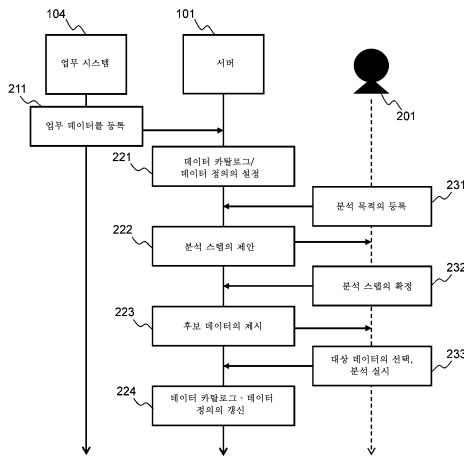
심사관 : 박미정

(54) 발명의 명칭 데이터의 이용 · 활용을 위한 데이터 준비를 지원하는 시스템, 및 그 방법

(57) 요약

복수의 업무 시스템으로부터 수집한 데이터의 이용 · 활용을 촉진하기 위해, 데이터의 이용 · 활용을 위한 데이터 준비를 지원하는 시스템, 및 그 방법을 제공한다.

본 발명의 시스템은, 분석 목적에 대해 권장되는 분석 스텝을 결정하고, 권장되는 분석 스텝에 이용 가능한 데이터의 조합을, 대상 데이터로부터 추출하고, 추출된 데이터의 조합의 복수의 각각을 평가하고, 평가의 결과에 기초하여, 데이터의 조합의 복수를 각각 순위 부여한 리스트를 작성하여, 유저가 소정의 데이터의 조합을 선택할 수 있도록 하였다.

대 표 도 - 도2

(52) CPC특허분류

G06F 16/283 (2019.01)

G06Q 10/10 (2021.08)

(56) 선행기술조사문현

JP2019106031 A

JP2015026188 A

JP2013077124 A

JP08077010 A

JP2019128646 A

JP2016004525 A

JP2010205218 A

JP2005190402 A

*는 심사관에 의하여 인용된 문현

명세서

청구범위

청구항 1

데이터의 이용·활용을 위한 데이터 준비를 지원하는 시스템으로서,

처리 장치와,

기억 장치를 구비하고,

상기 처리 장치는, 상기 기억 장치에 기록된 프로그램을 실행함으로써,

복수의 업무 시스템의 각각으로부터 업무 데이터를 수집하고, 당해 업무 데이터를 상기 기억 장치에 대상 데이터로서 적어도 일시적으로 축적하고,

유저 단말기로부터 상기 대상 데이터에 대한 분석 목적을 수신하고,

상기 분석 목적과 상기 대상 데이터에 기초하여, 상기 분석 목적에 대해 권장되는 분석 스텝을 결정하고,

상기 권장되는 분석 스텝에 이용 가능한 데이터의 조합을, 상기 대상 데이터로부터 추출하고,

상기 추출된 데이터의 조합의 복수의 각각을 평가하고,

당해 평가의 결과에 기초하여, 상기 데이터의 조합의 복수를 각각 순위 부여한 리스트를 작성하고,

당해 리스트를 상기 유저 단말기에 출력시키고,

유저가 상기 리스트로부터 소정의 데이터의 조합을 선택할 수 있도록 하고,

상기 기억 장치는, 상기 대상 데이터에 대한 기준 정보를 구비하고,

상기 처리 장치는,

상기 권장되는 분석 스텝을 상기 기준 정보에 기초하여 결정하는 것과,

상기 권장되는 분석 스텝을 결정하는 것을, 상기 대상 데이터에 대한 분석의 실시 단계의 복수 타입 중에서 어느 타입의 분석의 실시 단계에 결정함으로써 실행하는 것을 갖고,

상기 처리 장치는,

상기 분석 목적이, 상기 대상 데이터, 또는 당해 대상 데이터 및 상기 기준 정보에 적합한 정도를 산출하고,

상기 기준 정보의 상기 분석 목적에 대한 유효성의 정도를 산출하고,

양쪽의 산출 결과에 기초하여, 상기 어느 타입의 분석의 실시 단계를 결정하고,

상기 분석 목적은 KPI를 포함하고,

상기 기준 정보는 상기 대상 데이터에 대한 사전 데이터를 포함하고,

상기 처리 장치는,

상기 분석 목적이 상기 대상 데이터에 적합한 정도를 산출하는 것을, 상기 KPI와 상기 대상 데이터의 칼럼명을 비교함으로써 행하고,

상기 분석 목적이 상기 기준 정보에 적합한 정도를 산출하는 것을, 상기 KPI와 상기 사전 데이터를 비교함으로써 행하는,

상기 시스템.

청구항 2

제1항에 있어서,

상기 분석 목적은, 상기 복수의 타입의 분석의 실시 단계 중, 유저가 원하는 분석의 실시 단계를 포함하고,
상기 처리 장치는,

상기 양쪽의 산출 결과에 기초하여, 당해 유저가 원하는 분석의 실시 단계를 평가하고, 이 평가는 당해 분석의
실시 단계를 유지하거나, 또는 다른 분석의 실시 단계로 변경하는 것을 포함하고,

상기 평가의 결과에 기초하여, 상기 유저에게, 상기 권장되는 분석의 실시 단계를 제시하는,
시스템.

청구항 3

데이터의 이용 · 활용을 위한 데이터 준비를 지원하는 시스템으로서,

처리 장치와,

기억 장치를 구비하고,

상기 처리 장치는, 상기 기억 장치에 기록된 프로그램을 실행함으로써,

복수의 업무 시스템의 각각으로부터 업무 데이터를 수집하고, 당해 업무 데이터를 상기 기억 장치에 대상 데이
터로서 적어도 일시적으로 축적하고,

유저 단말기로부터 상기 대상 데이터에 대한 분석 목적을 수신하고,

상기 분석 목적과 상기 대상 데이터에 기초하여, 상기 분석 목적에 대해 권장되는 분석 스텝을 결정하고,

상기 권장되는 분석 스텝에 이용 가능한 데이터의 조합을, 상기 대상 데이터로부터 추출하고,

상기 추출된 데이터의 조합의 복수의 각각을 평가하고,

당해 평가의 결과에 기초하여, 상기 데이터의 조합의 복수를 각각 순위 부여한 리스트를 작성하고,

당해 리스트를 상기 유저 단말기에 출력시키고,

유저가 상기 리스트로부터 소정의 데이터의 조합을 선택할 수 있도록 하고,

상기 기억 장치는, 상기 대상 데이터에 대한 기준 정보를 구비하고,

상기 처리 장치는,

상기 권장되는 분석 스텝에 이용 가능한 데이터의 조합을, 상기 대상 데이터로부터 추출하는 것을,

상기 대상 데이터로부터 상기 분석 목적에 해당되는 데이터를 추출하는 것과, 그리고

상기 대상 데이터로부터 상기 기준 정보에 해당되는 데이터를 추출하는 것으로부터 실행하고,

상기 분석 목적에 해당되는 데이터와, 상기 기준 정보에 해당되는 데이터에 기초하여, 상기 데이터의 조합의 복
수를 구성하고,

당해 데이터의 조합의 복수의 각각을, 데이터양에 기초하여 평가하고,

상기 처리 장치는,

상기 대상 데이터로부터 상기 분석 목적에 해당되는 데이터를 추출하는 것을, 상기 분석 목적에 해당되는 칼럼
을 상기 대상 데이터로부터 추출하는 것으로부터 실행하고,

상기 대상 데이터로부터 상기 기준 정보에 해당되는 데이터를 추출하는 것을, 상기 대상 데이터로부터 상기 기
준 정보에 해당되는 칼럼을 상기 대상 데이터로부터 추출하는 것으로부터 실행하고,

상기 데이터의 조합으로서, 상기 분석 목적에 해당되는 칼럼과 상기 기준 정보에 해당되는 칼럼의 조합을 복수
구성하고,

상기 데이터양으로서, 당해 복수의 조합 각각의 레코드수를 산출하는,

시스템.

청구항 4

제1항에 있어서,
상기 처리 장치는,
상기 분석의 실시 단계의 복수 타입마다, 상기 추출된 데이터의 조합의 복수의 각각을 평가하는,
시스템.

청구항 5

제3항에 있어서,
상기 처리 장치는,
상기 복수의 조합의 각각에 대해, 복수 있는 데이터값 변동 패턴 중 어느 데이터값 변동 패턴을 구비할지를 판정하고,
데이터값 변동 패턴마다 상기 레코드수를 산출하는,
시스템.

청구항 6

제1항에 있어서,
상기 처리 장치는,
상기 대상 데이터의 칼럼을 상기 사전 데이터의 4W 정보의 각각에 기초하여 추출하는,
시스템.

청구항 7

제1항에 있어서,
유저에 의한 분석 목적 정보의 등록에 대한, 권장하는 분석 스텝 또는, 작업 항목에 관한 정보, 또한 각 분석 스텝에 있어서의 유저가 지정하는 KPI나 착안 데이터 항목과 관련된 이용 가능한 데이터의 조합에 관한 정보를, 유저에게 제시하기 위한 출력 장치를 갖는, 시스템.

청구항 8

데이터의 이용 · 활용을 위한 데이터 준비를 지원하는 방법이며,
컴퓨터는,
복수의 업무 시스템의 각각으로부터 업무 데이터를 수집하고, 당해 업무 데이터를 기억 장치에 대상 데이터로서 적어도 일시적으로 축적하고,
유저 단말기로부터 상기 대상 데이터에 대한 분석 목적을 수신하고,
상기 분석 목적과 상기 대상 데이터에 기초하여, 상기 분석 목적에 대해 권장되는 분석 스텝을 결정하고,
상기 권장되는 분석 스텝에 이용 가능한 데이터의 조합을, 상기 대상 데이터로부터 추출하고,
상기 추출된 데이터의 조합의 복수의 각각을 평가하고,
당해 평가의 결과에 기초하여, 상기 데이터의 조합의 복수를 각각 순위 부여한 리스트를 작성하고,
당해 리스트를 상기 유저 단말기에 출력시키고,
유저가 상기 리스트로부터 소정의 데이터의 조합을 선택할 수 있도록 하고,
상기 기억 장치는, 상기 대상 데이터에 대한 기준 정보를 구비하고,
상기 컴퓨터는,

상기 권장되는 분석 스텝을 상기 기준 정보에 기초하여 결정하는 것과,

상기 권장되는 분석 스텝을 결정하는 것을, 상기 대상 데이터에 대한 분석의 실시 단계의 복수 타입 중에서 어느 타입의 분석의 실시 단계에 결정함으로써 실행하는 것을 갖고,

상기 컴퓨터는,

상기 분석 목적이, 상기 대상 데이터, 또는 당해 대상 데이터 및 상기 기준 정보에 적합한 정도를 산출하고,

상기 기준 정보의 상기 분석 목적에 대한 유효성의 정도를 산출하고,

양쪽의 산출 결과에 기초하여, 상기 어느 타입의 분석의 실시 단계를 결정하고,

상기 분석 목적은 KPI를 포함하고,

상기 기준 정보는 상기 대상 데이터에 대한 사전 데이터를 포함하고,

상기 컴퓨터는,

상기 분석 목적이 상기 대상 데이터에 적합한 정도를 산출하는 것을, 상기 KPI와 상기 대상 데이터의 칼럼명을 비교함으로써 행하고,

상기 분석 목적이 상기 기준 정보에 적합한 정도를 산출하는 것을, 상기 KPI와 상기 사전 데이터를 비교함으로써 행하는,

상기 방법.

청구항 9

데이터의 이용 · 활용을 위한 데이터 준비를 지원하는 방법이며,

컴퓨터는,

복수의 업무 시스템의 각각으로부터 업무 데이터를 수집하고, 당해 업무 데이터를 기억 장치에 대상 데이터로서 적어도 일시적으로 축적하고,

유저 단말기로부터 상기 대상 데이터에 대한 분석 목적을 수신하고,

상기 분석 목적과 상기 대상 데이터에 기초하여, 상기 분석 목적에 대해 권장되는 분석 스텝을 결정하고,

상기 권장되는 분석 스텝에 이용 가능한 데이터의 조합을, 상기 대상 데이터로부터 추출하고,

상기 추출된 데이터의 조합의 복수의 각각을 평가하고,

당해 평가의 결과에 기초하여, 상기 데이터의 조합의 복수를 각각 순위 부여한 리스트를 작성하고,

당해 리스트를 상기 유저 단말기에 출력시키고,

유저가 상기 리스트로부터 소정의 데이터의 조합을 선택할 수 있도록 하고,

상기 기억 장치는, 상기 대상 데이터에 대한 기준 정보를 구비하고,

상기 컴퓨터는,

상기 권장되는 분석 스텝에 이용 가능한 데이터의 조합을, 상기 대상 데이터로부터 추출하는 것을,

상기 대상 데이터로부터 상기 분석 목적에 해당되는 데이터를 추출하는 것과, 그리고

상기 대상 데이터로부터 상기 기준 정보에 해당되는 데이터를 추출하는 것으로부터 실행하고,

상기 분석 목적에 해당되는 데이터와, 상기 기준 정보에 해당되는 데이터에 기초하여, 상기 데이터의 조합의 복수를 구성하고,

당해 데이터의 조합의 복수의 각각을, 데이터양에 기초하여 평가하고,

상기 컴퓨터는,

상기 대상 데이터로부터 상기 분석 목적에 해당되는 데이터를 추출하는 것을, 상기 분석 목적에 해당되는 칼럼

을 상기 대상 데이터로부터 추출하는 것으로부터 실행하고,

상기 대상 데이터로부터 상기 기준 정보에 해당되는 데이터를 추출하는 것을, 상기 대상 데이터로부터 상기 기준 정보에 해당되는 칼럼을 상기 대상 데이터로부터 추출하는 것으로부터 실행하고,

상기 데이터의 조합으로서, 상기 분석 목적에 해당되는 칼럼과 상기 기준 정보에 해당되는 칼럼의 조합을 복수 구성하고,

상기 데이터양으로서, 당해 복수의 조합 각각의 레코드수를 산출하는,

상기 방법.

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

발명의 설명

기술 분야

[0001] 본 발명은, 데이터의 이용·활용을 지원하기 위한 시스템, 특히 데이터의 이용·활용을 위한 데이터 준비를 지원하는 시스템에 관한 것으로, 예를 들어 복수의 업무 시스템이 갖는 다종, 대량의 데이터 중에서, 데이터 분석의 목적을 따른 데이터를 선출, 추출하거나 하기에 적합한 시스템에 관한 것이다.

배경 기술

[0002] 요즘, 기업 내에 축적된 다종 대량의 데이터를 이용하여 다양한 업무 과제를 개선하는 일이 행해지고 있다. 예를 들어, 일본 특허 공개 제2004-29971호 공보에는, 생산 공정에 있어서의 생산 장치나 처리 시각의 데이터 중에서 원하는 데이터 해석에 필요한 데이터만을 용이하게 추출하여, 수율 향상에 유효한 해석 결과를 얻기 위해, 데이터의 설명 변수에 대해 데이터 항목의 카테고리를 식별하는 부가 문자열을 부가하고, 데이터의 이상값을 특정값으로 치환 혹은 삭제하는 데이터 클렌징을 행하고, 데이터의 목적 변수의 변동에 기초하는 특징 정보를 얻어, 해석 처리 시에 데이터의 카테고리를 인식하고 카테고리에 대응한 조건 설정 및 해석 수순에 따라 데이터 해석을 효율적으로 자동 실행하는 데이터 해석 방법이 제안되어 있다.

[0003] 그리고 일본 특허 공개 제2016-181150호 공보에는, 요구되는 분석의 변경에 대응 가능한 데이터 처리를 행하기 위해, 입력된 모든 입력 데이터를 저장하는 데이터 웨어하우스와, 입력 데이터를 통합하여 통합 데이터를 생성한 후, 통합 데이터를 저장하는 통합 레이어와, 통합 데이터를, 불가산 항목의 하나 이상의 조합마다, 적어도 가산 항목의 수량 또는 불가산 항목의 수를 집계하여 복수의 집계 데이터를 생성한 후, 복수의 집계 데이터를 저장하는 집계 레이어와, 설정부에서 설정된 분석 데이터의 생성에 필요한 조건에 기초하여, 복수의 집계 데이터로부터 하나의 집계 데이터를 선택하고, 다시 하나의 집계 데이터로부터 분석 데이터를 추출한 후, 분석 데이터를 저장하는 분석 레이어를 갖는 데이터 처리 시스템이 제안되어 있다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 일본 특허 공개 제2016-181150호 공보

발명의 내용

해결하려는 과제

[0005]

근년, 교통, 전력, 산업 등의 많은 분야에 있어서, 다양화, 다각화되는 여러 가지 문제를 해결하기 위해, 부서나 업무를 획단하여 수집된 업무 데이터를 활용할 것이 요구되고 있다. 한편, 각각 제작기 구축, 운용되어 온, 복수의 업무 시스템마다 다른 각종, 대량의 데이터를 취급하기 위해, 유저에게는 데이터의 이해나 업무에 대한 경험이나 지식이 필요하기는 하지만, 그 레벨은 사람에 따라 달라, 그것이, 데이터를 사용한 과제의 분석을 진행하는 데 있어서 방해가 되고 있다. 예를 들어, 어느 유저는, 자신이 속하는 부서나 자신의 업무에 관한 데이터를 숙지하고 있어도, 다른 부서나 다른 업무에 관한 데이터를 파악하지 못한 경우가 많아, 그 결과, 복수의 부서나 업무에 걸쳐, 유저가 원하는 분석에 필요한 데이터를 선출하거나, 데이터를 가공하거나 하는 등, 분석을 위한 준비는 곤란해진다.

[0006]

그래서 복수의 업무마다의 차이, 업무마다의 데이터나 시스템의 차이에 대한 이해나 지식이 유저에게 부족해도, 유저가, 원활하게 데이터의 분석에 착수할 수 있도록, 데이터의 준비를 위한 작업 부하를 저감시킬 것이 요망된다.

[0007]

상기 특허문헌 1에 개시된 발명은, 유저가 대상 데이터를 이해하고 있는 것을 전제로 하는 것이며, 그 때문에, 분석 목적, 분석 수단, 및 분석을 위해 준비해야 할 데이터의 내용을, 유저가 사전에 정리해 두는 것이 필요하고, 특정 종류의 데이터에 대해, 유저가 상정한 목적하에서, 당해 데이터를 활용할 수 있는 것에 불과하다. 한편, 특허문헌 2에 개시된 발명에 활용할 수 있는 데이터는, 통합 데이터가 될 수 있는 데이터로 한정된다. 복수의 업무 시스템으로부터의 각종 다양한 데이터를 목적에 맞게 통합할 수 있다고는 할 수 없다. 또한 통합 데이터, 집계 데이터로부터 목적에 맞는 분석 데이터를 작성하기 위해서는, 유저는, 원래의 데이터를 모두 이해하고 있을 필요가 있다.

[0008]

본 발명은, 복수의 업무 시스템으로부터 수집한 데이터의 이용·활용을 촉진시키기 위해, 데이터의 이용·활용을 위한 데이터 준비를 지원하는 시스템, 및 그 방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0009]

상기 목적을 해결하기 위해, 본 발명은, 데이터의 이용·활용을 위한 데이터 준비를 지원하는 시스템으로서, 처리 장치와, 기억 장치를 구비하고, 상기 처리 장치는, 상기 기억 장치에 기록된 프로그램을 실행함으로써, 복수의 업무 시스템의 각각으로부터 업무 데이터를 수집하고, 당해 업무 데이터를 상기 기억 장치에 대상 데이터로서 적어도 일시적으로 축적하고, 유저 단말기로부터 상기 대상 데이터에 대한 분석 목적을 수신하고, 상기 분석 목적과 상기 대상 데이터에 기초하여, 상기 분석 목적에 대해 권장되는 분석 스텝을 결정하고, 상기 권장되는 분석 스텝에 이용 가능한 데이터의 조합을, 상기 대상 데이터로부터 추출하고, 상기 추출된 데이터의 조합의 복수의 각각을 평가하고, 당해 평가의 결과에 기초하여, 상기 데이터의 조합의 복수를 각각 순위 부여한 리스트를 작성하고, 당해 리스트를 상기 유저 단말기에 출력시켜, 유저가 상기 리스트로부터 소정의 데이터의 조합을 선택할 수 있도록 하였다.

발명의 효과

[0010]

본 발명에 따르면, 복수의 업무 시스템의 각종 다양, 대량의 데이터에 기초하여, 과제 해결 등의 분석 등을 위해, 이들 데이터의 이용·활용을 진행할 때의 데이터 준비 작업에 있어서의 유저의 부하를 경감시킴으로써, 데이터의 이용·활용을 신속하면서 고품질로 진행할 수 있다.

도면의 간단한 설명

[0011]

도 1은 데이터의 이용·활용을 위한 데이터 준비를 지원하는 시스템 구성의 일례를 나타내는 블록도이다.

도 2는 서버가 데이터의 이용·활용을 위한 데이터 준비를 지원하는 동작을 실현하기 위한 시퀀스의 일례이다.

도 3은 서버의 모듈 구성의 일례이다.

도 4는 유저가 작성하는 분석 목적 정보, 서버에서 유지하는 기준 정보의 구성의 일례이다.

도 5a는 분석 스텝으로서 전체 상 파악을 유저가 지정하면, 서버가 작성하는, 데이터 준비 테이블의 일례이다.

도 5b는 분석 스텝으로서, "특이 사상 추출"을 유저가 지정하면, 서버가 작성하는, 데이터 준비 테이블의 일례이다.

도 5c는 분석 스텝으로서, "요인 분석/예측"을 유저가 지정하면, 서버가 작성하는, 데이터 준비 테이블의 일례이다.

도 5d는 서버로부터, 유저에게 제시, 또는 제안되는, 데이터의 조합을 관리하는 관리 테이블의 일례를 나타낸다.

도 6은 서버가 유저에게 분석 스텝을 제안하기 위한 동작의 일례에 관한, 흐름도이다.

도 7a는 유저가 등록한 분석 스텝의 적부를 판정하기 위한 매트릭스의 제1 예이다.

도 7b는 유저가 등록한 분석 스텝의 적부를 판정하기 위한 매트릭스의 제2 예이다.

도 7c는 유저가 등록한 분석 스텝의 적부를 판정하기 위한 매트릭스의 제3 예이다.

도 8은 서버가 유저에게, 분석 스텝으로서의 전체 상 파악에 이용 가능한 데이터의 조합을, 유저에게 제시하기 위한 동작의 일례를 나타내는 흐름도이다.

도 9는 서버가, 분석 스텝으로서의 특이 사상 추출에 이용 가능한 데이터의 조합을, 유저에게 제시하기 위한 동작의 일례를 나타내는 흐름도이다.

도 10은 서버가, 분석 스텝으로서의 요인 분석/예측에 이용 가능한 데이터의 조합을, 유저에게 제시하기 위한 동작의 일례를 나타내는 흐름도이다.

도 11a는 서버로부터 유저 단말기에 대해 제공된, 데이터 이용·활용을 지원하기 위한 그래피컬 유저 인터페이스의 일례이다.

도 11b는 서버로부터 유저 단말기에 대해 제공된, 데이터 이용·활용을 지원하기 위한 그래피컬 유저 인터페이스의 다른 예이다.

도 11c는 서버로부터 유저 단말기에 대해 제공된, 데이터 이용·활용을 지원하기 위한 그래피컬 유저 인터페이스의 또 다른 예이다.

도 12는 도 8의 흐름도에 있어서, 복수의 칼럼의 조합을 설명하기 위한 테이블이다.

도 13은 도 10의 흐름도에 있어서, 복수의 칼럼의 조합을 설명하기 위한 테이블이다.

발명을 실시하기 위한 구체적인 내용

[0012] 다음으로, 본 발명의 실시 형태에 대해 설명한다. 도 1은, 본 발명을 실현하기 위한 계산기 시스템의 일례의 구성도이다. 이 계산기 시스템은, 복수의 업무 시스템으로부터 수집한 데이터의 이용·활용을 위한 데이터 준비를 지원하는 서버(101)(컴퓨터)를 구비한다. 서버(101)는, 시스템을 이용하는 유저가 입력한, 데이터 분석의 목적과, 서버(101)가 수집한, 복수의 업무 시스템의 데이터를 대조하여, 유저에 대해, 최적의 분석 스텝을 제안하고, 아울러, 이 분석 스텝에 이용 가능한 유용도가 높은 데이터를 제안함으로써, 데이터의 이용·활용을 위한 데이터 준비를 지원한다.

[0013] 유저는, 시스템에 분석 목적을 등록하는 것만으로, 그리고 시스템으로부터 제안, 또는 제공된 데이터에 기초하여, 과제 해결 등을 위한 분석을 실행함으로써, 유저 자신이 데이터를 수집하거나, 수많은 데이터 중에서 유용한 데이터를 추출하는 등의 데이터 준비에 관한 작업을 강요받는 일 없이, 유저가 의도한, 데이터 분석을 행할 수 있다.

[0014] 서버(101)에는, 복수의 업무 시스템(104, 105, 106)과, 복수의 유저 단말기(102, 103)가 접속되어 있다. 유저 단말기는, 유저가 희망하는 분석 목적을 서버(101)에 등록하고, 이어서 분석 목적을 실현하는 데 있어서의 분석 스텝, 및 이 분석 스텝에 이용 가능한 데이터의 제안을, 서버(101)로부터 받고, 이어서 이들을 확인 후, 데이터 분석을 실시하거나, 혹은 서버(101)에 대해 데이터 분석을 실시시킬 수 있다.

[0015] 복수의 업무 시스템의 각각으로부터, 서버(101)에, 유저가 원하는 데이터 분석의 대상으로서, 서버(101)가 참조하는 데이터(이하, 대상 데이터라고 함)가 제공된다. 또한, 복수의 업무 시스템의 각각으로부터 서버(101)로의 데이터 제공은 네트워크(108)를 거친 통신에 의해 실시하는 경우뿐만 아니라, 네트워크(108)를 거치지 않고, 예를 들어 사람의 손을 거친 서버(101)로의 데이터 저장을 행하는 것이어도 된다. 복수의 업무 시스템으로서는, 철도 분야를 예로 들면, 예를 들어 열차의 운행 관리 시스템, 역의 관리 시스템, 그리고 보수 관리 시스템을 예

시할 수 있다.

[0016] 업무 시스템으로부터 서버(101)로 보내지는 데이터는, 특별히 제한되지 않아도 되고, 예를 들어 데이터 테이블, 그 밖에 바이너리 데이터로서의 센서 정보 등, 데이터의 종류, 데이터의 방식 등이 제한되지 않는다. 유저는, 복수의 업무 시스템을 획단하여, 소정의 과제나 문제점의 원인을 종합적으로 찾기 위해, 서버(101)에, 데이터 분석의 목적을 등록한다. 이 목적으로서는, 수송업을 예로 들면, 예를 들어 "열차의 지연"이 있고, 즉, 지연의 원인을, 복수의 업무 시스템을 획단하여 특정함으로써, 지연의 예방, 억제에 유용하게 사용하려고 하는 경우가 있다.

[0017] 서버(101)와, 복수의 유저 단말기(102, 103, ·····)의 각각은 네트워크(107)를 통해 접속되어 있다. 서버(101)와, 복수의 업무 시스템(104, 105, 106, ·····)의 각각은 네트워크(108)를 통해 상호 접속되어 있다.

[0018] 데이터 이용·활용, 및 데이터의 준비를 위한 기반으로서의 서버(101)의 주된 하드웨어 구성은, 기억 장치(메인 메모리, 하드 디스크, 외부 스토리지 시스템의 기억 영역)(111), 처리 장치(CPU)(112), 그리고 통신 장치(113)이다. 유저 단말기(102, 103)도 마찬가지이다. 처리 장치(112)는, 비일시적 기록 매체(기억 장치)(111)에 대상 데이터를 적어도 일시적으로 기록한다.

[0019] 도 2는, 서버(101)가 데이터의 이용·활용을 위한 데이터 준비를 지원하는 동작을 실현하기 위한 시퀀스의 일례이다. 또한, 유저(201)는, 복수의 부서를 획단하여 수집된 다양한 데이터에 대해, 소정의 분석 스텝에 기초하여, 문제 발견, 해결책의 입안 등을 행하려고 하는 사람, 또는 그것을 위한 애플리케이션이면 된다.

[0020] 복수의 업무 시스템(104-106)의 각각은, 서버(101)에 업무 데이터를 등록한다(211). 서버(101)는, 업무 데이터의 기억 영역에 액세스하여, 업무 데이터를 참조하고, 데이터 카탈로그, 데이터 정의를 설정한다(221). 서버(101)가 데이터 카탈로그 등을 설정한다는 것은, 데이터 카탈로그 등을 작성하는 것, 또는 개신하는 것을 포함한다. 서버(101)는, 후술하는 바와 같이, 유저가 입력한 분석 목적과, 자신이 작성한 기준 정보(데이터 카탈로그, 데이터 정의)에 기초하여, 분석 양태(분석 스텝, 후보 데이터 리스트)를 작성하여, 이것을 유저에게 제시한다.

[0021] 데이터 카탈로그, 그리고 데이터 정의 등의 기준 정보는, 서버(101)가, 유저(201)가 의도하는 분석 목적과, 업무 데이터에 기초하여, 유저(201)에게, 최적의 분석 스텝을 제안하고, 그리고 이 분석 스텝에 이용 가능한 유용도가 높은 데이터를 제안하기 위해, 업무 데이터를 분별하고, 평가하는 등, 업무 데이터의 해석을 위한 기준, 또는 지표 등의 일례로서 이해되면 된다. 데이터 카탈로그, 데이터 정의의 상세는 후술한다.

[0022] 유저(201)는, 유저 단말기(102)를 통해, 원하는, 데이터 분석의 목적을 서버(101)에 등록한다(231). 서버(101)는, 업무 시스템 데이터, 그리고 기준 정보(데이터 카탈로그, 데이터 정의)에 기초하여, 유저가 등록한 분석 목적을 평가하여, 유저에게 권장하는 분석 스텝을 판정, 판별, 결정, 그리고 선정하거나 하여 확정하고, 이것을 유저 단말기(102)에 보고한다(222).

[0023] 유저(201)는, 서버(101)로부터 제시되고, 권장된 분석 스텝을 확인하거나, 또는 복수의 분석 스텝 중에서 소정의 분석 스텝을 선택하거나 하여, 분석 스텝을 확정시켜, 이것을 서버(101)에 보고한다(232).

[0024] 서버(101)는, 복수의 업무 시스템으로부터 수집한 데이터 중에서, 분석 스텝에 적합한, 복수의 후보 데이터를 추출하여, 이것을 유저(201)에게 제시한다(223). 유저(201)는, 서버(101)로부터 송신된, 복수의 후보 데이터의 리스트를 참조하여, 분석 대상으로 하는 데이터를 결정, 혹은 선택 등에 의해 설정하여, 확정한 분석 스텝(232)에 기초하여, 분석 대상 데이터를 분석한다. 또한, 이것을 서버(101)에 보고한다(233). 서버(101)는, 스텝 222, 스텝 223, 그리고 스텝 233의 결과에 기초하여, 기준 정보(데이터 카탈로그/데이터 정의)를 개신한다(224).

[0025] 데이터 카탈로그, 데이터 정의 등의 기준 정보는, 서버(101)의 관리자, 또는 운영자에 의해 작성, 또는 개신되어도 된다. 서버(101)는, 유저의 분석 목적, 업무 데이터에 기초하는 기계 학습을 행함으로써, 데이터 카탈로그 등을 작성, 또는 개신해도 된다.

[0026] 분석 목적(후술하는 도 4: 401)은 데이터 분석의 타깃을 단적으로 나타내는 정보이면 되고, 예를 들어 "열차 지연"이다. 유저(201)는, 분석 목적을, 예를 들어 "분석 스텝", "KPI", 그리고 "착안 데이터 항목"에 의해 기술하여 서버에 등록할 수 있다. "분석 스텝", 그리고 "KPI"는 필수 정보로 하고, "착안 데이터 항목"은 임의 정보로 해도 된다. KPI는, 데이터의 분석에 있어서, 유저가 주목하는 요소이며, 예를 들어 앞서 서술한 "열차 지연"이다. KPI는, 키워드 외에, 산출식에 의해 정의되어도 된다.

- [0027] 착안 데이터 항목은, 예를 들어 KPI를 평가하기 위한 축으로서 착안되어야 할 요소, 또는 KPI와 동시에 감시해야 할 명칭이다. 예를 들어, KPI가 "열차 지연"으로서, 착안 데이터 항목은, "발차 시각", "도착 시각", "운행 일시", 그리고 "지연 시분" 중 적어도 하나이다.
- [0028] 분석 스텝은, 업무상의 문제 해결 등을 위해 분석을 진행하는 데 있어서의 실시 단계를 나타낸 것으로, 복수의 타입이 있고, 주로, 대상 데이터의 전체 상을 파악하기 위한 양태, 대상 데이터로부터 특이 사상을 추출하기 위한 양태, 그리고 대상 데이터로부터 요인을 분석하여, 결과를 예측하는 양태가 예시된다. 이들을, 이후, (1) 전체 상 파악, (2) 특이 사상 추출, (3) 요인 분석/예측이라고 하기로 한다.
- [0029] 유저(201)는, 복수의 분석 스텝을 분석 목적에 포함시킬 수 있다. 복수의 스텝에는, 우선도, 순번 등의 우열의 구별이 있어도 된다.
- [0030] 상기 분석 스텝의 각각의 양태에 대해, 유저(201)가 사용할 수 있는 계산 방법으로서, 예를 들어 전체 상 파악에는 경향 분석, 특이 사상 추출에는 이상점 검출, 및/또는 변화점 추출, 요인 분석/예측에는, 회귀 분석/중회귀 분석, 및/또는 분류/클러스터 분석이 있다.
- [0031] 서버(101)는, 기억 장치(111)에, 기준 정보("정의 정보"라고 칭해도 됨)로서, 데이터 카탈로그, 데이터 정의(221)를 갖는다. 데이터 카탈로그(도 4: 402)에는, 서버(101)가 복수의 업무 시스템(104)으로부터 수집한 대상 데이터의 데이터 카탈로그와, 대상 데이터를 가공한 가공 데이터의 데이터 카탈로그가 있다. 가공 데이터의 데이터 카탈로그에는, 대상 데이터로부터 가공 데이터를 얻기 위한 산출식에 관한 정보를 포함해도 된다.
- [0032] 대상 데이터에 관한 카탈로그로서, 예를 들어 "칼럼명", "명칭 · 의미", "보충 정보" 등의 항목으로 이루어지는 것이 있다. 일례로서, {arrival time, 도착 시각, . . .}, {KRT, 킬로정, . . .}이 있다. 또한, 칼럼명이란, 대상 데이터(데이터 테이블)의 칼럼에 기록된 명칭이다.
- [0033] 가공 데이터에 관한 카탈로그로서, "데이터명", "산출식", "명칭 · 의미", "보충 정보" 등의 항목으로 이루어지는 것이 있고, 일례로서, {delay, "실적 시각"- "계획 시각", 지연 시분, . . .}이 있다.
- [0034] 앞서 서술한 데이터 정의(221)는, 예를 들어 4W 사전(도 4: 403)과, 데이터 추이 패턴(도 4: 404)을 구비한다. 4W 사전은, 대상 데이터의 전체 상의 경향을 4W(When, Where, What, Who)의 관점에서 파악하기 위한 것이며, 4W의 카테고리마다 키워드를 포함한다. 예를 들어, 4W 사전 데이터는, "키워드", "카테고리"의 조합으로 이루어지고, 일례로서 {키워드, 카테고리}={시각, when}, {date, when}, {킬로정, where}, {Kilometrage, where}, {역 코드, where}이다.
- [0035] 데이터 추이 패턴은, 4W 사전에 값의 추이를 추가로 포함하는 것이며, 예를 들어 "데이터명 · 종별", "카테고리", "데이터형", "값 변화 폭", "하한", 그리고 "상한"으로 이루어진다. 일례로서, {킬로정, where, integer, 1, 0, 300}, {도착 시각(분), when, integer, 5, 0, 1440}이 있다.
- [0036] 도 3은, 서버(101)의 모듈 구성의 일례이다. 서버(101)에는, 앞서 서술한 도 2에 있어서 설명한, 여러 기능을 실현하기 위한 미들웨어(301)가 실장되어 있다. 미들웨어(301)는, 서버(101)의 기억 장치(111)에 보존된 프로그램을 실행하는 처리 장치(112)에 의해 실현되는, 복수의 기능 모듈을 구비하고 있다. 그리고 미들웨어(301)는, 업무 시스템(104)의 대상 데이터(311), 대상 데이터(311)로부터 특징량으로서 추출되거나 하여 가공된 가공 데이터(312)를 기억 장치(111)에 구비하고 있다.
- [0037] 미들웨어(301)는, 대상 데이터(311), 가공 데이터(312)를 관리하는 데이터 관리 모듈(321)을 구비하고, 데이터 카탈로그(402), 데이터 정의(315), 그리고 유저(201)가 등록한 분석 목적의 적부를 판정하기 위한 매트릭스(313)를 기억 장치(111)에 구비하고 있다.
- [0038] 미들웨어(301)는, 앞서 서술한 기준 정보(데이터 카탈로그, 데이터 정의)를 관리하는 기준 정보 관리 모듈(322), 서버(101)에 액세스하여 데이터 준비에 관한 작업을 행하는 유저를 관리하는 유저 관리 모듈(323), 정의 정보를 참조하고, 그리고 유저의 분석 목적 정보(401)에 기초하여, 유저에게 분석 스텝, 및/또는 작업 항목의 제안을 행하는 분석 스텝 제안 모듈(324)을 구비한다.
- [0039] 미들웨어(301)는, 데이터 정의(315)를 참조하여, 유저의 분석 목적 정보(401)에 기초하여 분석용 데이터를 제안하는 분석용 데이터 제안 모듈(325)과, 유저에 대한 제안 정보를 데이터 준비 테이블(314)로서 갱신, 관리하는 제안 실행 관리 모듈(326)과, 유저 단말기(102, 103)에, 미들웨어(301)의 기능에 액세스하기 위한 인터페이스를 제공하는 인터페이스 모듈(327)과, 네트워크(107, 108)를 통해, 유저 단말기(102, 103), 그리고 업무 시스템(104 . . .)과 통신을 행하는 데이터 통신 모듈(328)과, 분석 스텝이 전체 상 파악인 경우의 분석용 데이터

제안을 위한, 예를 들어 경향 분석 등의 처리를 실행하는 전체 상 파악 모듈(331)과, 분석 스텝이 특이 사상인 경우의 분석용 데이터 제안을 위한, 예를 들어 특이점(예를 들어, 이상점, 큰 변화점, 큰 바이어스)을 추출하는 처리를 실행하는 특이 사상 추출 모듈(332)과, 분석 스텝이 요인 분석/예측인 경우의 분석용 데이터 제안을 위한, 예를 들어 회귀 분석, 중회귀 분석 등의 처리를 실행하는 요인 분석·예측 모듈(333)을 구비한다.

- [0040] 앞서 서술한 기준 정보 관리 모듈(322)은, 기준 정보(데이터 카탈로그(402), 데이터 정의(315), 매트릭스(313))를 관리한다. 예를 들어, 이 정보는, 산업 분야마다(철도, 전력 등) 존재해도 되고, 기준 정보 관리 모듈(322)은, 산업 분야마다 기준 정보를 관리하고, 업무 데이터, 또는 분석 목적 등에 따라서, 소정 분야의 기준 정보를 선택하도록 해도 된다.
- [0041] 또한, 앞서 서술한 모듈이란, 처리 장치가 프로그램을 실행함으로써 실현되는 기능이며, 예를 들어 부, 수단, 요소, 회로, 또는 유닛으로 환언되어도 된다.
- [0042] 다음으로, 서버(101)에 의해 작성되고, 유저에게 제시되는 분석용 데이터의 후보 리스트에 대해 설명한다. 도 5a의 501은, 분석 스텝으로서 "전체 상 파악"을 유저가 지정하면, 서버(101)가 작성하는, 상기 데이터 준비 테이블(314)(후보 리스트)의 일례이다. 분석용 데이터 제안 모듈(325)이, 이 테이블을 작성, 또는 생성한다. 이 테이블은, 순위(511), 식별 정보(512), KPI(513), 착안 데이터 1(When)(514), 착안 데이터 2(Where)(515), 착안 데이터 3(What)(516), 착안 데이터 4(Who)(517), 레코드수(518), 그리고 출력 파일(519)을 구비한다.
- [0043] 순위(511)에는, 식별 정보(512)에 의해 특정되는 데이터 조합(행)에 대해, 우선도에 기초하여 할당되는 순위에 관한 정보가 저장된다. 데이터 조합이란, 대상 데이터(311)의 테이블이 "KPI 데이터"를 칼럼명으로서 포함하는, 당해 칼럼의 레코드와, "when"에 해당되는 착안 데이터 1(예를 들어, 앞서 서술한 "시각")을, 대상 데이터(311)의 테이블이 칼럼명으로서 포함하는, 당해 칼럼의 레코드와, "where"에 해당되는 착안 데이터 2(예를 들어, 앞서 서술한 "킬로정")에 관한 마찬가지의 레코드와, "what"에 해당되는 착안 데이터 3(예를 들어, 앞서 서술한 "열변")에 관한 마찬가지의 레코드와, 그리고 "who"에 해당되는 착안 데이터 4(예를 들어, "운전사 ID")에 관한 마찬가지의 레코드의 조합을 말한다.
- [0044] 또한, 착안 데이터 1-4(514-517) 각각에는, 분석 목적 정보(401)를 구성하는 유저 입력의, "착안 데이터 항목"과 동일 또는 유사한 것이 있으면, 이것이 기재된다. 착안 데이터 1-4(514-517) 중, 해당되는 것이 없는, 즉, 4W 사전에 등록이 없고, 또한 착안 데이터 항목도 규정되어 있지 않은 경우에는, 공백이 기록된다.
- [0045] KPI(513)에는, 분석 목적 정보(401)를 구성하는, 유저 입력의 "KPI"와, 동일, 또는 유사한 것이 저장된다.
- [0046] 레코드수(518)에는, 식별 정보(512)에 의해 특정되는 데이터 조합에 있어서의, 유효 상태에 있는 레코드수가 저장된다. 출력 파일(519)에는, 식별 정보(512)에 의해 특정되는 데이터 조합의 레코드를 출력하는 파일의 패스가 저장된다.
- [0047] 도 5b의 502는, 분석 스텝으로서, "특이 사상 추출"을 유저가 지정하면, 서버(101)가 작성하는, 데이터 준비 테이블(314)의 일례이다. 특이 사상 리스트이기도 한 테이블(502)에 있어서, 테이블(501)과 동일한 부호에 대한 설명은 동일하다. 테이블(502)은, 특이점(이상점) 수(520), 특이점(변화점) 수(521), 특이점(바이어스) 수(522)를 구비한다.
- [0048] 특이점(이상점) 수(520)에는, 식별 정보(512)에 의해 특정되는 데이터 조합에 있어서의, KPI로부터 추출한 특이점의 종별이 "이상점"인 것의 수에 관한 정보가 저장된다.
- [0049] 특이점(변화점) 수(521)에는, 식별 정보(512)에 의해 특정되는 데이터 조합에 있어서의, KPI로부터 추출한 특이점의 종별이 "큰 변화점"인 것의 수에 관한 정보가 저장된다. 특이점(바이어스) 수(522)에는, 식별 정보(512)에 의해 특정되는 데이터 조합에 있어서의 KPI로부터 추출한 특이점의 종별이 "큰 바이어스"인 것의 수에 관한 정보가 저장된다.
- [0050] 도 5c의 503은, 분석 스텝으로서, "요인 분석/예측"을 유저가 지정하면, 서버(101)가 작성하는, 데이터 준비 테이블(314)의 일례이다. 목적 변수·설명 변수 조합 리스트이기도 한 테이블(503)에 있어서, 테이블(501)과 동일한 부호에 대한 설명은 동일하다.
- [0051] 목적 변수·설명 변수 조합 리스트(503)는, 목적 변수(530), 설명 변수(531), 레코드수(518), 판정값 1(533), 판정값 2(534), 판정값 3(535), 판정값 4(536)를 구비한다. 목적 변수(530)에는, 분석 목적에 포함되는 KPI가 저장된다. 설명 변수 데이터(531)에는, KPI명을 포함하는 칼럼 이외에, 유저에 의해 선택된, 칼럼의 명칭이 기

록된다.

- [0052] 판정값 1(533)에는, 식별 정보(512)에 의해 특정되는, 데이터의 조합의 순위(511)를 산출하기 위해 사용하는 첫 번째의 판정값 데이터가 저장되고, 판정값 2(534)에는, 두 번째의 판정값 데이터가 저장되고, 판정값 3(535)에는, 세 번째의 판정값 데이터에 관한 정보가 저장되고, 판정값 4(536)에는, 네 번째의 판정값 데이터에 관한 정보가 저장된다. 이를 판정값에 대해서는 도 10에 있어서, 상세하게 설명한다.
- [0053] 도 5d는, 서버(101)로부터, 유저에게 제시, 또는 제안되는, 데이터의 조합을 관리하는 관리 테이블(504)의 일례이다. 이 테이블은, 데이터 관리 모듈(321)에 의해 작성, 또는 생성된다. 관리 테이블(504)은, 서버(101)가, 유저로부터의 분석 목적 정보(401)와, 기준 정보(221)와, 그리고 대상 데이터(311)에 기초하여 유저에게 제안하는 데이터에 관한 정보를 저장한다.
- [0054] 제안 실행 관리 모듈(326)은, 유저로부터의 분석 목적 정보(401)를 접수하여, 분석 스텝을 권장하고, 그리고 분석에 이용되는 데이터를 제안할 때, 테이블(504)을 작성, 또는 생성한다. 이 테이블(504)은, 식별 정보(541), KPI(513), 데이터 조합(543), 이용 횟수(544), 이용 인원수(545), 생성 일시(546)를 포함한다.
- [0055] 식별 정보(541)에는, 유저에게 데이터 제안을 식별하기 위한 정보가 저장된다. KPI(513)에는, 식별 정보(541)에 의해 특정되는 데이터 제안에 있어서의 KPI 데이터가 저장된다. 데이터 조합(543)에는, 식별 정보(541)에 의해 특정되는 유저에게 제안되는 데이터의 조합이, 분석 스텝이 전체 상 파악, 또는 특이 사상 추출인 경우에 저장되고, 분석 스텝이 요인 분석·예측인 경우, 설명 변수의 조합이 저장된다.
- [0056] 이용 횟수(544)는, 식별 정보(541)에 의해 특정되는 데이터 제안에 있어서의 데이터 조합의 이용 횟수에 관한 정보가 저장되고, 유저에 의한 이용 시마다 생성된다.
- [0057] 이용 인원수(545)에는, 식별 정보(541)에 의해 특정되는 데이터 제안에 있어서의 데이터 조합의 이용 인원수에 관한 정보가 저장된다. 이 이용 인원수는, 총인원수여도 되고, 다른 유저의 인원수여도 된다. 이 정보는, 식별 정보(541)에 의해 특정되는 데이터 제안이, 유저에게 이용될 때마다 생성된다. 생성 일시(546)에는, 후보 리스트(501)의 레코드가 생성된 일시가 저장된다.
- [0058] 도 6은, 서버(101)가 유저에게 분석 스텝을 권장하기 위한 동작의 일례에 관한 흐름도이다. 서버(101)는, 유저의 입력 정보(231)와, 기준 정보(221), 대상 데이터(311)에 기초하여, 유저의 분석 목적이 적절한지 여부를 판정하여, 판정 결과를, 유저에게 권장하는 분석 스텝으로서 출력한다.
- [0059] 서버(101)는, 유저로부터 분석 목적 정보(401)의 등록(231)이 있으면, 흐름도를 개시한다. 분석 스텝 제안 모듈(324)은, 스텝 601에 있어서, 유저로부터의 분석 목적 정보(401)에 포함되는, KPI와 착안 데이터 항목 각각의 명칭을, 대상 데이터(311)의 테이블의 칼럼명과 비교하여, 비교 결과를, 일치도로서 산출한다. 동일한 명칭이 "일치"가 되는 것은 물론이지만, 유사한 명칭을 "일치"에 포함시켜도 된다. "유사"란, 예를 들어 유의어를 말한다. "일치도"란, 일치의 누계, 일치의 비율, 그리고 역치와의 비교 등이면 된다. 모듈(324)은, 대상 데이터가 복수 있는 경우에는, 대상 데이터마다 일치도를 산출하고, 예를 들어 대상 데이터마다의 일치도를 누계하여, 복수의 대상 데이터의 일치도로 해도 된다.
- [0060] 이어서, 분석 스텝 제안 모듈(324)은, 스텝 602에 있어서, KPI, 그리고 착안 데이터 항목 각각의 명칭과, 데이터 카탈로그(402)의 데이터 항목명을 비교하여 앞서 서술한 일치도를 계산한다.
- [0061] 이어서, 분석 스텝 제안 모듈(324)은, 스텝 603에 있어서, KPI, 그리고 착안 데이터 항목 각각의 명칭과, 데이터 정의(315)(4W 사전, 데이터 추이 패턴 사전)에 있는 데이터 항목명을 비교하여, 일치도를 구한다.
- [0062] 스텝 601에 있어서, 일치도가 높다고 하는 것은, 유저의 분석 목적에 대한, 대상 데이터의 데이터양이 부족하지 않은 것, 및/또는 대상 데이터의 질이 높은 것을 나타내는 것이며, 그리고 스텝 602, 603에 있어서, 일치도가 높다고 하는 것은, 대상 데이터를 분별, 분류, 판별, 또는 평가하거나 하기 위한 정보가, 유저의 분석 목적에 적합한 것을 나타낸다. 일치도를, 예를 들어 적합도, 친화도, 해당도 등으로 환연해도 된다.
- [0063] 분석 스텝 제안 모듈(324)은, 스텝 604에 있어서, 유저로부터의 분석 목적 정보(401)에 있는 분석 스텝이, "요인 분석/예측"인지 여부를 판정한다. 분석 스텝 제안 모듈(324)이, 스텝 604를 부정 판정하면, 스텝 605에 있어서, 기준 정보(데이터 카탈로그(402), 데이터 정의(315))의 충실도를 산출한다. 충실도라 함은, 기준 정보의 분석 목적에 대한 유효성, 유용성, 또는 신뢰성을 나타내는 지표, 예를 들어 정보의 풍부함의 정도를 나타내는 지표이며, 이것은 예를 들어 데이터 카탈로그(402), 데이터 정의(315)에 대한 정보 등록 건수, 그리고 참조 횟

수 등의 액세스 빈도에 기초하여 결정되어도 된다. 충실도를 유효도 등으로 환언해도 된다.

- [0064] 분석 스텝 제안 모듈(324)이 스텝 604를 긍정 판정하면, 스텝 606으로 이동하여, 분석 스텝으로서의 "요인 분석/예측"에 있어서의, 설명 변수 후보가 되는 가공 데이터(312)의 충실도를 산출한다(스텝 606). 이 충실도는, 가공 데이터의 칼럼명의 수, 유효한 레코드의 수의 많음으로부터 산출되어도 된다.
- [0065] 분석 스텝 제안 모듈(324)은, 스텝 607에 있어서, 스텝 605 내지 606의 결과에 기초하여, 도 7a 등에 나타내는, 유저가 등록한 분석 스텝 231의 적부를 판정하기 위한 매트릭스(313)를 참조하여, 유저에게 권장해야 할, 분석 스텝, 및/또는 작업 항목을, 판정, 판별하거나 함으로써 결정한다.
- [0066] 분석 스텝 제안 모듈(324)은, 스텝 607의 결정 내용을 스텝 608에 있어서, 유저에게 제시하여, 유저의 확인, 선택 등을 요구한다. 이상에 의해, 서버(101)는, 흐름도를 종료한다. 또한, 분석 스텝, 그리고 작업 항목을 통합하여, 예를 들어 분석 양태, 또는 분석 방법 등이라고 칭해도 된다.
- [0067] 분석 스텝 제안 모듈(324)은, 유저에게 제안해야 할 분석 스텝을 판정하기 위한 기준인 매트릭스로서, 유저가 분석 목적 정보(분석 스텝)로서, "전체 상 파악"을 선택한 경우에는, 매트릭스(701)(도 7a)를 채용하고, "특이 사상 추출"을 선택한 경우에는, 매트릭스(702)(도 7b)를 채용하고, "요인 분석/예측"을 선택한 경우에는, 매트릭스(703)(도 7c)를 채용한다.
- [0068] 판정 매트릭스(701)(도 7a)는, 스텝 601-603의 분석 목적 정보와의 일치도와, 스텝 605의 데이터 카탈로그, 데이터 정의의 충실도와의 상관을 규정한 것이며, 분석 스텝 제안 모듈(324)이, 일치도, 그리고 충실도를 각각 소정의 역치와 비교하여, 그 높고 낮음을 결정해도 된다.
- [0069] 판정 매트릭스(701)의 "분석 목적 정보와의 일치도"는, 스텝 601-603 각각의 일치도를, 예를 들어 가산한 것, 평균한 것 등이면 된다. 스텝 601-603 각각의 일치도 중, 소정의 스텝의 일치도를 우선시키도록 해도 된다. 판정 매트릭스(701)의 "데이터 카탈로그/데이터 정의의 충실도"는, 각각의 충실도, 예를 들어 가산한 것, 평균한 것 등이면 된다. 한쪽의 충실도를 우선시키도록 해도 된다.
- [0070] 일치도가 "높고" 또한 충실도가 "높은" 경우, 그리고 일치도가 "높고" 또한 충실도가 "낮은" 경우, 유저의 분석 목적에 적합한, 대상 데이터는 충분량 존재할 수 있으므로, 분석 스텝 제안 모듈(324)은, 유저가 서버(101)에 등록한 분석 스텝으로서의 "전체 상 파악"을, 그대로 권장하여 유저에게 제시한다.
- [0071] 한편, 일치도가 "낮고" 또한 충실도가 "높은" 경우, 대상 데이터양이 부족할 가능성이 있기 때문에, "전체 상 파악"이 즉시 실시되는 것은 바람직하지 않으므로, 분석 스텝 제안 모듈(324)은, "전체 상 파악"을 유지하면서도, 그것을 실현하는 데 필요한 작업 항목으로서, 우선은, "데이터 추가"를 권장한다.
- [0072] 또한, 일치도가 "낮고" 또한 충실도가 "낮은" 경우, 유저의 분석 목적에 적합한 데이터는 존재하고 있었다고 해도, 해당초, 데이터 카탈로그(402), 그리고 데이터 정의(315)의 질이 충분하지 않을 가능성이 있으므로, 분석 스텝 제안 모듈(324)은 분석 스텝으로서 "전체 상 파악"의 실시는 바람직하지 않다고 판정하여, 작업 항목으로서, 통계에 의한 데이터 이해 촉진, 데이터 카탈로그(402), 데이터 정의(315)의 확충을 권장한다.
- [0073] 판정 매트릭스(702)(도 7b)의 "분석 목적 정보와의 일치도", "높음", "낮음"의 의의는, 판정 매트릭스(701)의 것과 동일하다. "분석 목적 정보에 일치하는 데이터 충실도"는, 스텝 601-603에서 유저로부터의 분석 목적 정보(401)와의 일치도가 높다고 판정되는, 대상 데이터(311)에 있어서의 데이터, 또는 데이터 카탈로그(402), 데이터 정의(315)의 데이터 항목에 해당되는 대상 데이터(311)(또는, 가공 데이터(312))에 있어서의 데이터의 충실도를 나타낸다.
- [0074] 이 충실도는, 이들 데이터(데이터)의 유효한 레코드의 수의 많음으로부터 산출되어도 된다. 일치도가, "높고" 또한 충실도가 "높은" 경우, 분석 스텝 제안 모듈(324)은, 유저의 분석 목적에 적합한, 대상 데이터양은 충분할 수 있으므로, 유저가 등록한, 분석 스텝으로서의 "특이 사상 추출"을, 그대로 권장한다.
- [0075] 한편, 일치도가 "높고" 또한 충실도가 "낮은" 경우, 유저의 분석 목적에 적합한 데이터가 충분히 존재할 수 있어도, 특이 사상으로서의 데이터 레코드수가 상대적으로 부족할 가능성이 있으므로, 분석 스텝 제안 모듈(324)은 분석 스텝으로서 "특이 사상 추출"의 실시는 곤란하다고 판정하여, 작업 항목으로서, 대상 데이터를 추가할 것을 권장한다.
- [0076] 또한, 일치도가 "낮고" 또한 데이터의 충실도가 "높은" 경우, 또는 일치도가 "낮고" 또한 충실도가 "낮은" 경우, 분석 스텝 제안 모듈(324)은, 유저가 요망하는 분석 목적에 적합한, 대상 데이터가 부족하거나, 혹은 유

저가, 충분히, 데이터를 이해하지 못했을 가능성이 있다고 판단하여, 분석 스텝으로서 "특이 사상 추출"을 실시하는 것은 곤란하여, 유저에게, 우선, "전체 상 파악"으로 변경하여, "전체 상 파악"으로 돌아가, 이제부터 데이터 분석을 개시할 것을 권장한다. 이 경우, 분석 스텝 제안 모듈(324)은, 판정 매트릭스(701)(도 7a)에 기초하여, "전체 상 파악"으로서의 적부를 판정한다.

[0077] 판정 매트릭스(703)(도 7c)의 "분석 목적 정보와의 일치도"에 대해, "높음", "낮음"의 의의는, 판정 매트릭스(701)의 것과 동일하다. 판정 매트릭스(703)의 "가공 데이터의 충실도"는, 가공 데이터에 대한 것이며, "충실도", 그리고 그 "높음", "낮음" 자체의 의의는, 앞서 설명한 바와 같다.

[0078] 일치도가 "높고" 또한 충실도가 "높은" 경우, 유저의 분석 목적에 적합한 대상 데이터양은 충분하므로, 분석 스텝 제안 모듈(324)은, 유저가 지정한 분석 스텝으로서의 "요인 분석/예측"을 그대로 권장한다.

[0079] 일치도가 "높고" 또한 충실도가 "낮은" 경우, 분석 스텝 제안 모듈(324)은 유저의 분석 목적에 적합한, 대상 데이터가 충분히 존재할 수 있어도, 요인 분석/예측을 실시하려면 설명 변수가 되는, 가공 데이터의 레코드수가 부족할 수 있으므로, 분석 스텝으로서의 "요인 분석/예측"의 실시는 곤란하다고 판정하여, 작업 항목으로서 가공 데이터의 확충을 유저에게 권장한다.

[0080] 일치도가 "낮고" 또한 충실도가 "높은" 경우, 분석 스텝 제안 모듈(324)은, 유저의 분석 목적에 적합한, 대상 데이터 자체가 부족할 가능성이 있으므로, 유저에게, "전체 상 파악"으로 돌아가 검토할 것을 권장한다.

[0081] 일치도가 "낮고" 또한 충실도가 "낮은" 경우, 분석 스텝 제안 모듈(324)은, 유저의 분석 목적에 적합한, 대상 데이터 자체가 부족할 가능성이 있거나, 혹은 데이터 이해가 충분하지는 않을 가능성이 있기 때문에, 또한 요인 분석/예측을 실시하려면 설명 변수가 되는 데이터의 레코드수가 부족할 가능성이 있기 때문에, 유저에게, "전체 상 파악"으로 돌아갈 것, 그리고 작업 항목으로서 가공 데이터를 확충할 것을 권장한다.

[0082] 도 8은, 서버(101)가 유저에게, 분석 스텝으로서의 "전체 상 파악"에 이용 가능한 데이터의 조합을, 유저에게 제시하기 위한 동작의 일례를 나타내는 흐름도이다. 서버(101)는, 판정 매트릭스(701)에 기초하여, 분석 스텝으로서, "전체 상 파악"을 결정하면, 도 8의 흐름도를 개시시킨다.

[0083] 분석용 데이터 제안 모듈(325)은, 스텝 801에 있어서, 유저(201)로부터 접수한 분석 목적 정보(401(231))로부터 KPI와 착안 데이터 항목을 추출한다. 이어서, 분석용 데이터 제안 모듈(325)은, 스텝 802에 있어서, 대상 데이터(311)의 테이블로부터 상기 KPI에 해당되는 칼럼, 즉, KPI의 명칭과 동일 또는 유사명의 칼럼을 추출한다. 도 12에 있어서, 대상 데이터(311)는, 2개의 데이터(테이블)(311A, 311B)를 구비하고, 테이블(311A)의 칼럼 B가 KPI와 동일명의 칼럼이다.

[0084] 분석용 데이터 제안 모듈(325)은, 스텝 803에 있어서, 대상 데이터(311)로부터 착안 데이터 항목에 해당되는 칼럼을 추출한다. 분석용 데이터 제안 모듈(325)은, 스텝 804에 있어서, 대상 데이터(311)로부터, 4W 사전(221)의 4W(When, Where, What, Who) 각각의 키워드에 해당되는(키워드와 동일, 또는 유사한), 대상 데이터(311)의 칼럼을 추출한다. 또한, 분석용 데이터 제안 모듈(325)은, 스텝 802, 및/또는 스텝 803에서 추출된 칼럼이, 스텝 804에서 다시 추출된 경우에는, 이것을 제외한다.

[0085] 분석용 데이터 제안 모듈(325)은, 스텝 805에 있어서, 대상 데이터(311)에 있어서의 각 칼럼의 레코드를 참조하여, 데이터 추이 패턴 정의(404)에 있는 4W 사전(221)의 4W의 각각의 키워드의 데이터값의 추이 패턴에 해당되는 칼럼을 대상 데이터(311)로부터 추출한다. 또한, 분석용 데이터 제안 모듈(325)은, 스텝 802, 및/또는 스텝 803에서 추출된 칼럼이, 스텝 805에서 다시 추출된 경우에는, 이것을 제외한다.

[0086] 분석용 데이터 제안 모듈(325)은, 스텝 806에 있어서, 스텝 803, 스텝 804, 또는 스텝 805에서 추출된 칼럼의 우선도를 설정한다. 분석용 데이터 제안 모듈(325)은, 예를 들어 상기 세 스텝에서 추출된 칼럼의 우선도를 두 스텝 혹은 한 스텝에서만 추출된 칼럼보다 우선도를 높게 한다.

[0087] 분석용 데이터 제안 모듈(325)은, 스텝 807에 있어서, 스텝 804-805의 칼럼(4W 후보 칼럼)의 리스트를, 스텝 806에 관한 우선도를 알 수 있도록, 유저에게 제시하고, 스텝 808에서 유저로부터의 요구가 있으면, 4W 후보 칼럼의 축소를 실시하여, 일부의 칼럼을 데이터 분석 대상으로부터 제외할 수 있다.

[0088] 분석용 데이터 제안 모듈(325)은, 스텝 809에 있어서, 스텝 802에서 추출한, KPI에 해당되는 칼럼(KPI 칼럼)과, 스텝 803에서 추출한 착안 데이터 항목에 해당되는 칼럼(착안 데이터 항목 칼럼), 스텝 804-808에 관한 칼럼(4W 후보 칼럼)으로 이루어지는 조합을 작성한다. 이후, 이 조합을, "합성 칼럼 조합"이라고 한다.

- [0089] 스텝 803에서 추출한 착안 데이터 항목에 해당되는 칼럼과, 스텝 804-808에 관한 4W 후보 칼럼을 통합하여, KPI 칼럼에 대해 축 후보가 되는 "축 후보 칼럼"이라고 칭하는 것으로 한다.
- [0090] 도 12에 있어서, 테이블(311A)의 칼럼 C가 축 후보 칼럼 1이고, 테이블(311A)의 칼럼 D가 축 후보 칼럼 2이고, 테이블(311B)의 칼럼 G가 축 후보 칼럼 2이고, 테이블(311B)의 칼럼 H가 축 후보 칼럼 3이다(칼럼 D와 칼럼 G의 명칭은 동일 또는 유사). 테이블(311C)가 합성 칼럼 조합이다.
- [0091] 분석용 데이터 제안 모듈(325)은, 동일 또는 유사한 명칭의 KPI 칼럼이 복수의 데이터 테이블에서 중복되는 경우, 그리고 동일 또는 유사한 명칭의 축 후보 칼럼이 복수의 데이터 테이블에서 중복되는 경우, 중복되는 칼럼을 교환하도록 하여 합성 칼럼 조합을 복수 작성한다. 도 12의 예를 근거로 하여 설명하면, 분석용 데이터 제안 모듈(325)은, 합성 칼럼 조합으로서의 테이블(311C)을, 축 후보 칼럼 2가 테이블(311A)의 칼럼 C인 것과, 축 후보 칼럼 2가 테이블(311B)의 칼럼 G인 것, 두 개 작성한다. 즉, 분석용 데이터 제안 모듈(325)은, 합성 칼럼 조합 1(칼럼 B, C, D, H)과 합성 칼럼 조합 2(칼럼 B, C, G, H)의 두 개를 정의한다. 동일명의 KPI 칼럼의 수를 "k"라고 하고, 동일명의 축 후보 1의 수를 "m1", 동일명의 축 후보 2의 수를 "m2"… 동일명의 축 후보 n의 수를 "mn"이라고 하면, 합성 칼럼 조합의 수는 " $k*m1*m2*\dots*mn$ "이 된다. 혹은 동일 또는 유사한 명칭의 축 후보 칼럼끼리를 키로 하여 테이블(311A, 311B)을 결합함으로써 합성 칼럼 조합을 작성한다.
- [0092] 분석용 데이터 제안 모듈(325)은, 스텝 810에 있어서, 복수의 합성 칼럼 조합으로부터, 유효한 레코드수가 0인 칼럼 조합을 제외한다. 유효한 레코드수가 0인 칼럼 조합이란, 칼럼 조합의 모든 레코드에 데이터가 기록되어 있지 않은 것을 말한다.
- [0093] 이어서, 분석용 데이터 제안 모듈(325)은, 스텝 811에 있어서, 합성 칼럼 조합의 각각에 대해, 각각의 조합의 레코드의 특성, 속성, 예를 들어 레코드에 포함되는 복수의 데이터값의 변동 패턴을 판정한다. 분석용 데이터 제안 모듈(325)은, 이 판정 결과를, 분석용 데이터의 후보의 작성에 반영시킨다.
- [0094] 데이터값의 변동 패턴은, 예를 들어 "연속값", "연속 반복값", 또는, "이산값"이다. "연속값"은, 수치 데이터이며, 모든 레코드에서 거의 값이 다른 패턴이다. "연속 반복값"은, 수치 데이터이며, 모든 레코드에 대해, 데이터값이 일정 범위 내에서 증감하는 것을 반복하는 패턴이다. "이산값"은, 문자열 데이터, 또는 수치 데이터이며, 유니크한 값의 수가 레코드수의 일정 비율 이하인 패턴이다.
- [0095] 분석용 데이터 제안 모듈(325)은, 스텝 812에 있어서, 합성 칼럼 조합마다 데이터값 변동 패턴을 판별하여, 패턴이 "연속값"인 것을 판별하면, 스텝 813으로 이행하여, 합성 칼럼 조합의 각각에 대해, 모든 합성 칼럼 조합을 결합하여, 총 레코드의 수를 산출한다.
- [0096] 분석용 데이터 제안 모듈(325)은, 스텝 812에 있어서, 데이터값 변동 패턴을 판별하여, 패턴이 "연속 반복값"인 것을 판별하면, 스텝 814로 이행하여, 합성 칼럼 조합의 각각에 대해, 데이터의 반복의 범위마다 레코드를 분류하고, 또한 스텝 815로 이행하여, 반복 범위마다, 합성 칼럼 조합의 각각을 결합하여, 레코드수를 산출한다.
- [0097] 분석용 데이터 제안 모듈(325)은, 스텝 812에 있어서, 데이터값 변동 패턴을 판별하여, 패턴이 "이산값"인 것을 판별하면, 스텝 816으로 이행하여, 스텝 811에 있어서 선택한 축 후보 칼럼과 4W의 동 분류에서 "이산값"인 것이 판별된 다른 축 후보 칼럼을 하나 이상 선출하고, 스텝 817에 있어서, KPI 칼럼과 스텝 816에서 선출한 복수의 축 후보 칼럼을 결합하여, 문자열 데이터와 수치 데이터의 각 분류에서의 레코드수를 산출한다.
- [0098] 분석용 데이터 제안 모듈(325)은, 스텝 818에 있어서, 스텝 811-817의 결과에 기초하여, 레코드수를 데이터값 변동 패턴마다 집계하고, 스텝 819에 있어서, 합성 칼럼 조합 전부에 처리가 완료되어 있는지 여부를 판정하여, 이것을 부정하면, 스텝 811로 리턴하고, 이것을 긍정 판정하면, 스텝 820으로 이행한다.
- [0099] 분석용 데이터 제안 모듈(325)은, 스텝 820에 있어서, 데이터값 변동 패턴마다, 모든 조합의 각각의 유용도를, 예를 들어 유효 상태의 레코드수(데이터양)를 키로 하여 평가하고, 모든 조합을 평가값의 내림차순으로 정렬하여 후보 리스트(도 5a의 501)를 작성하여 유저에게 제시한다.
- [0100] 분석용 데이터 제안 모듈(325)은, 후보 리스트(501)의 순위(511)에, 스텝 820에서 결정된 순위를 저장하고, KPI, 그리고 착안 데이터 항목을, KPI(513), 착안 데이터 1-4(514-517)에, 유효 레코드수를 레코드수(518)에, 데이터 조합의 파일 출력처를 출력 파일(519)에 각각 저장한다.
- [0101] 분석용 데이터 제안 모듈(325)은, 스텝 821에 있어서, 후보 리스트(501)에 포함되는, 칼럼 조합 각각에 대해, 경향 분석을 실시하고, 그 결과를 그래프로서 작성하여, 유저에게 제시한다. 유저는, 그래프를 참조하여, 소정의 조합을 지정, 선택할 수 있다. 유저는, 후보 리스트 중에서 원하는 조합을 선택해도 된다. 이 경우, 선택

된 조합에 대해서만, 그래프가 작성되어도 된다.

- [0102] 분석용 데이터 제안 모듈(325)은, 스텝 822에 있어서, 유저에 의해 지정된 조합에 관한 칼럼의 레코드를 CSV 파일 등으로서 출력한다. CSV 파일은, 유저가 새롭게 대상 데이터의 경향 분석을 실시할 때에 사용할 수 있다.
- [0103] 분석용 데이터 제안 모듈(325)은, 스텝 823에 있어서, 스텝 820에서 작성된 후보 리스트(501), 스텝 822에서 지정된 조합에 기초하여, 데이터 카탈로그(402)와 데이터 정의(221)를 생성, 또는 추가한다. 따라서, 데이터 분석이 진행됨에 따라서, 데이터 카탈로그와 데이터 정의가 개량되어 간다.
- [0104] 도 9는, 서버(101)가, 분석 스텝으로서의 "특이 사상 추출"에 이용 가능한 데이터의 조합을, 유저에게 제시하기 위한 동작의 일례를 나타내는 흐름도이다. 서버(101)는, 판정 매트릭스(702)에 기초하여, 분석 스텝으로서, "특이 사상 추출"을 결정하면, 도 9의 흐름도를 개시시킨다.
- [0105] 도 9의 흐름도에 있어서, 도 8의 흐름도의 스텝 801-819의 공정은 동일하다. 분석용 데이터 제안 모듈(325)은, 스텝 903에 있어서, 모든 칼럼 조합의 각각에 대해 순위 부여를 행하여, 순위가 역치 이상인 조합을 추출한다. 순위 부여는, 복수의 조합 각각에 있어서의 유효한 레코드의 많음에 기초한 것이면 된다.
- [0106] 분석용 데이터 제안 모듈(325)은, 스텝 904에 있어서, 스텝 903에서 추출한, 복수의 조합 중 하나의 조합의 데이터에 대해 특이점의 유무를 판정하여, 스텝 905에서 특이점의 판정을 긍정하면, 스텝 906에 있어서, 특이점의 수를 산출한다. 특이점이란, 예를 들어, 이상점, 큰 변화점, 큰 바이어스 등이면 된다. 분석용 데이터 제안 모듈(325)은, 상기 특이점에 관한 정보를 특이 사상 리스트(502)에 추가한다.
- [0107] 분석용 데이터 제안 모듈(325)은, 스텝 905를 부정하면, 스텝 906을 거치는 일 없이, 스텝 907로 이행하고, 앞서 서술한 조합 전부에 대해, 스텝 904, 905, 906을 적용하였는지 여부를 판정하여, 부정 판정하면 스텝 904로 돌아가고, 긍정 판정하면 스텝 908로 이행한다.
- [0108] 분석용 데이터 제안 모듈(325)은, 스텝 908에 있어서, 모든 조합 각각의 유용도를, 예를 들어 특이점의 수에 기초하여, 오름차순으로 정렬하거나 하여 순위 부여하고, 특이 사상 리스트(502)를 작성하여, 이것을 유저에게 제시한다. 특이점의 수는, 예를 들어 이상점의 수, 변화점의 수, 그리고 바이어스의 수를 합계한 것, 혹은 평균이어도 된다.
- [0109] 분석용 데이터 제안 모듈(325)은, 특이 사상 리스트(502)의 순위(511)에, 조합마다의 순위를 저장하고, KPI를 KPI(513)에 저장하고, 착안 데이터 항목을 착안 데이터 1-4(514-517)에 저장하고, 조합마다의 유효 레코드수를 레코드수(518)에 저장하고, 조합마다의 특이점(이상점, 큰 변화점, 큰 바이어스)의 수를 특이점수(520-522)에 저장하고, 조합마다의 데이터 파일의 출력처를 출력 파일(519)에 저장한다.
- [0110] 분석용 데이터 제안 모듈(325)은, 스텝 909에 있어서, 모든 조합마다 특이 사상 산출 그래프를 작성하여, 유저에게 제시한다. 그 이후는 도 8의 흐름도의 스텝 822, 823과 동일하다.
- [0111] 도 10은, 서버(101)가, 분석 스텝으로서의 요인 분석/예측에 이용 가능한 데이터의 조합을, 도 8, 도 9의 흐름도와 마찬가지로 리스트로서, 유저에게 제시하기 위한 동작의 일례를 나타내는 흐름도이다. 서버(101)는, 판정 매트릭스(703)에 기초하여, 분석 스텝으로서, "요인 분석/예측"을 결정하면, 도 10의 흐름도를 개시시킨다.
- [0112] 분석용 데이터 제안 모듈(325)은, 도 10의 흐름도를 개시하면, 스텝 1001에 있어서, 유저로부터 접수한 분석 목적 정보(401)에 기초하여, KPI(KPI 명칭, KPI 산출식)를 추출한다.
- [0113] 분석용 데이터 제안 모듈(325)은, 스텝 1002에 있어서, KPI의 명칭에 해당되는, 대상 데이터(311), 및/또는 가공 데이터(312)의 칼럼명을 추출하고, 이것을 목적 변수 칼럼으로 한다. 분석용 데이터 제안 모듈(325)은, KPI 산출식에 기초하여, 목적 변수 칼럼을 작성해도 된다. 도 13에 있어서, 대상 데이터(311C)의 칼럼 B가, KPI에 해당되는 목적 변수 칼럼이다.
- [0114] 분석용 데이터 제안 모듈(325)은, 스텝 1003에 있어서, 대상 데이터(311)의 테이블, 또는 가공 데이터(312)의 테이블로부터 설명 변수의 후보가 되는 칼럼의 소정수를 선택한다. 설명 변수의 후보가 되는 칼럼은, 대상 데이터(311)의 테이블, 또는 가공 데이터(312)의 테이블의 칼럼 중, 스텝 1002에서 선출된 목적 변수 칼럼을 제외한 칼럼이다.
- [0115] 분석용 데이터 제안 모듈(325)은, 목적 변수 칼럼에 대한, 설명 변수 칼럼 중 하나 또는 복수의 조합을 모두 작성한다. 도 13은, 대상 데이터(311)(테이블(311A))의 칼럼 B 이외의 모든 칼럼, 가공 데이터(312)(테이블(311D))의 모든 칼럼이 설명 변수 칼럼이 될 수 있음을 나타내고 있다. 예를 들어, 선택할 설명 변수 칼럼의

수가 2개라고 하면, 분석용 데이터 제안 모듈(325)은, 대상 데이터(311A)의 칼럼 B 이외의 모든 칼럼과, 가공 데이터(311D)의 모든 칼럼으로부터 2개의 칼럼의 모든 조합을 복수 작성하고, 각각의 조합과 목적 변수 칼럼을 맞추어, 목적 변수 칼럼과 설명 변수 칼럼의 합성 칼럼으로 이루어지는 테이블을 복수 구성한다. 도 13의 311E-1 내지 311E-4 ······ 의 각각이 합성 칼럼으로 이루어지는 테이블이다.

[0116] 분석용 데이터 제안 모듈(325)은, 스텝 1004에 있어서, 대상 데이터의 테이블, 그리고 가공 데이터의 테이블의 모든 칼럼 사이의 상관 계수를 산출한다. 분석용 데이터 제안 모듈(325)은, 스텝 1005에 있어서, 스텝 1002에서 추출한 목적 변수 칼럼과, 스텝 1003에서 선출한 수의 설명 변수의 후보 칼럼의 모든 조합에 대해, 회귀 분석, 또는 중회귀 분석 등의 학습을 실시한다.

[0117] 분석용 데이터 제안 모듈(325)은, 스텝 1006에 있어서, 목적 변수 칼럼과 설명 변수 칼럼의 모든 조합의 각각에 대해 유용도를 연산하고, 연산 결과에 기초하여, 모든 조합 각각을 순위 부여한 리스트를 작성한다. 유저는 원하는 조합을 선택할 수 있다.

[0118] 복수의 조합 중 각각의 조합의 유용도는, 예를 들어 (1) 목적 변수와 설명 변수의 상관 계수의 절댓값(스텝 1004), (2) 스텝 1005에서의, 학습 결과의 이점(결정 변수, 정답률), (3) 목적 변수의 칼럼과 설명 변수의 칼럼에 있어서의, 유효한 값을 포함하는 레코드의 총 수, 그리고 (4) 설명 변수간의 상관 계수의 절댓값(스텝 1004) 중 적어도 하나의 우열이면 되고, 분석용 데이터 제안 모듈(325)은 유용도의 오름차순, 또는 내림차순에 의해, 조합을 순위 부여한다.

[0119] 분석용 데이터 제안 모듈(325)은, 유용도 (1)-유용도 (4) 각각의 순위를 종합한 순위를 결정하여, 이것을 목적 변수 · 설명 변수 조합 리스트(503)의 순위(511)에, KPI를 목적 변수(KPI)(530)에, 하나 이상의 설명 변수의 명칭을 설명 변수(531)에, 모든 조합의 유효 레코드수를 레코드수(518)에, 유용도 (1)-(4)를 유용도 1-4(533-536)에, 조합 각각의 레코드의 파일의 출력처를 출력 파일(519)에, 각각 저장한다.

[0120] 분석용 데이터 제안 모듈(325)은, 스텝 1007에 있어서, 스텝 1006에서 작성한 리스트에 포함되는 목적 변수 · 설명 변수의 조합마다 학습 결과 그래프를 작성하고 유저에게 제시하여, 유저가 원하는 조합을 선택할 수 있게 한다. 여기서 리스트의 상위로부터 지정된 수의 목적 변수 · 설명 변수의 조합만의 그래프를 작성하는 것도 가능하다.

[0121] 그리고 분석용 데이터 제안 모듈(325)은, 스텝 822에 있어서, 스텝 1007의 결과를 참조한 유저에 의해 지정된 설명 변수 칼럼 및 목적 변수 칼럼에 관한 데이터 레코드를 CSV 파일 등에 출력한다. 이 파일은, 유저가 스스로 요인 분석/예측을 실시할 때에 사용된다.

[0122] 분석용 데이터 제안 모듈(325)은, 스텝 1006에서의 리스트(503)의 작성 결과, 이 리스트로부터의 유저에 의한 선택의 결과를 기초로, 데이터 카탈로그, 데이터 정의(315)의 변경, 정보의 초기화 행한다.

[0123] 도 11a-도 11c는, 서버(101)로부터, 유저 단말기(102, 103)에 대해 제공된, 데이터 이용 · 활용을 지원하기 위한 그래픽 인터페이스의 예이다. 제안 실행 관리 모듈(326)은, 분석용 데이터 제안 모듈(325)과, 분석 방법 제안 모듈(324)에 기초하여, 분석 목적을 리퀘스트한 유저 단말기에, 그래픽 인터페이스용 화면을 표시시킨다.

[0124] 도 11a의 화면(1101)에 있어서, 유저가 지정한 분석 스텝으로서의 "전체 상 파악"과, 이것에 대한, 서버(101)가 "데이터 추가"를 권장하는 것이 제1 영역(1111)에 표시되어 있다.

[0125] 또한, 분석 목적 정보(401)와 대상 데이터의 일치도와, 데이터 카탈로그, 데이터 정의 정보의 충실도로 이루어지는 제2 영역(1112)과, 데이터 추가를 위한 부족 데이터와, 데이터 카탈로그의 저장처, 그리고 데이터 정의의 저장처로 이루어지는 제3 영역(1113)이 보충 정보로서 표시되어 있다.

[0126] 도 11b의 화면(1102)은, 권장 분석 스텝이 전체 상 파악인, 상기 후보 리스트(501)의 일례이다. 복수의 칼럼을 조합한 데이터에 대해, 그 복수(스텝 820)가 유용도순으로 나열되어 있다(1121). 영역 1122는, 데이터 조합마다, 경향 분석의 결과 얻어진 그래프의 예이다(스텝 821).

[0127] 도 11c의 화면(1103), 권장 분석 스텝이 요인 분석/예측인, 상기 목적 변수 · 설명 변수 조합 리스트(503)의 일례이다. 복수의 칼럼을 조합한 데이터에 대해, 그 복수(스텝 1006)가 유용도순으로 나열되어 있다(1131). 영역 1132는, 데이터 조합마다, 학습의 결과를 나타내는 그래프이다.

[0128] 앞서 서술한 실시 형태에 따르면, 서버(101)는, 유저(201)가 복수의 업무 시스템의 다종 다양, 대량의 데이터를 활용하여 과제 분석 등을 행할 때, 유저가 분석 목적을 서버에 등록하면, 유저의 부하를 경감시키면서, 분석에

유용한 데이터의 리스트를 유저에게 제시할 수 있다. 유저는, 데이터 이용 · 활용을 행하기 전의 데이터의 준비 작업을 신속하면서 용이하게 실현할 수 있게 된다.

[0129]

서버(101)는, 유저에게 권장한 분석 스텝을 실행하려고 하는 당해 유저에 대해, 분석 목적과 관련된 데이터를 제시하려고 할 때, 분석 목적(KPI, 착안 데이터 항목)과 관련된 칼럼과, 분석 목적과 적합성이 있고, 또한 충실판도 높은 기준 정보와 관련된 칼럼을 조합하여, 양 칼럼의 레코드를 유저에게 제시하므로, 유저는, 분석 목적을 서버에 등록하기만 하면, 분석 목적과 관련된 데이터의 조합을 광범위하게 취득할 수 있다.

부호의 설명

[0130]

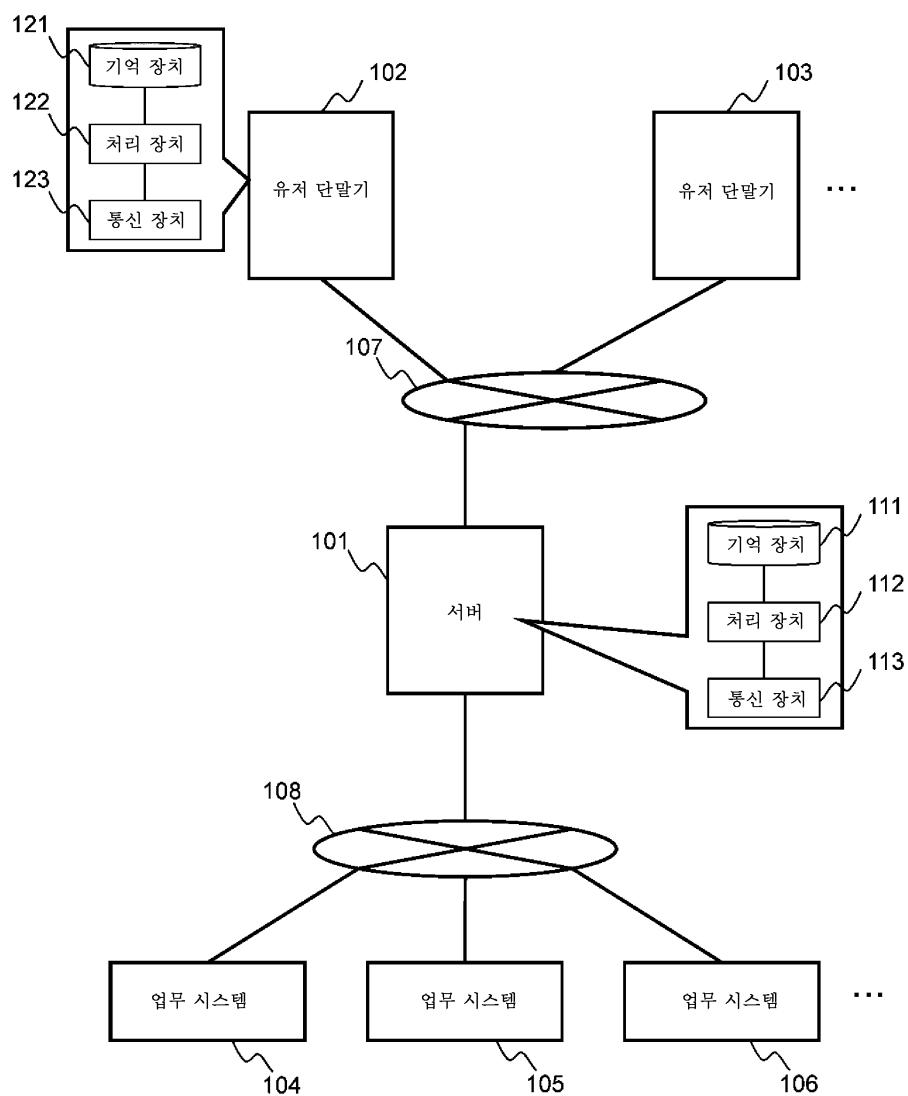
101: 서버

102, 103: 유저 단말기

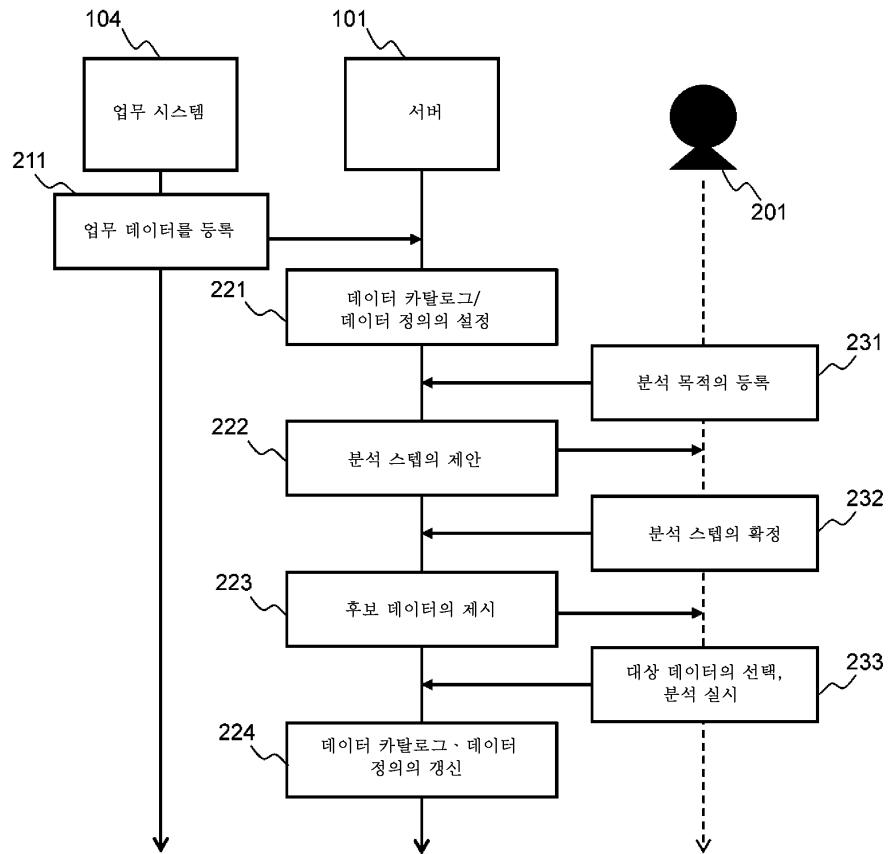
104-106: 업무 시스템

도면

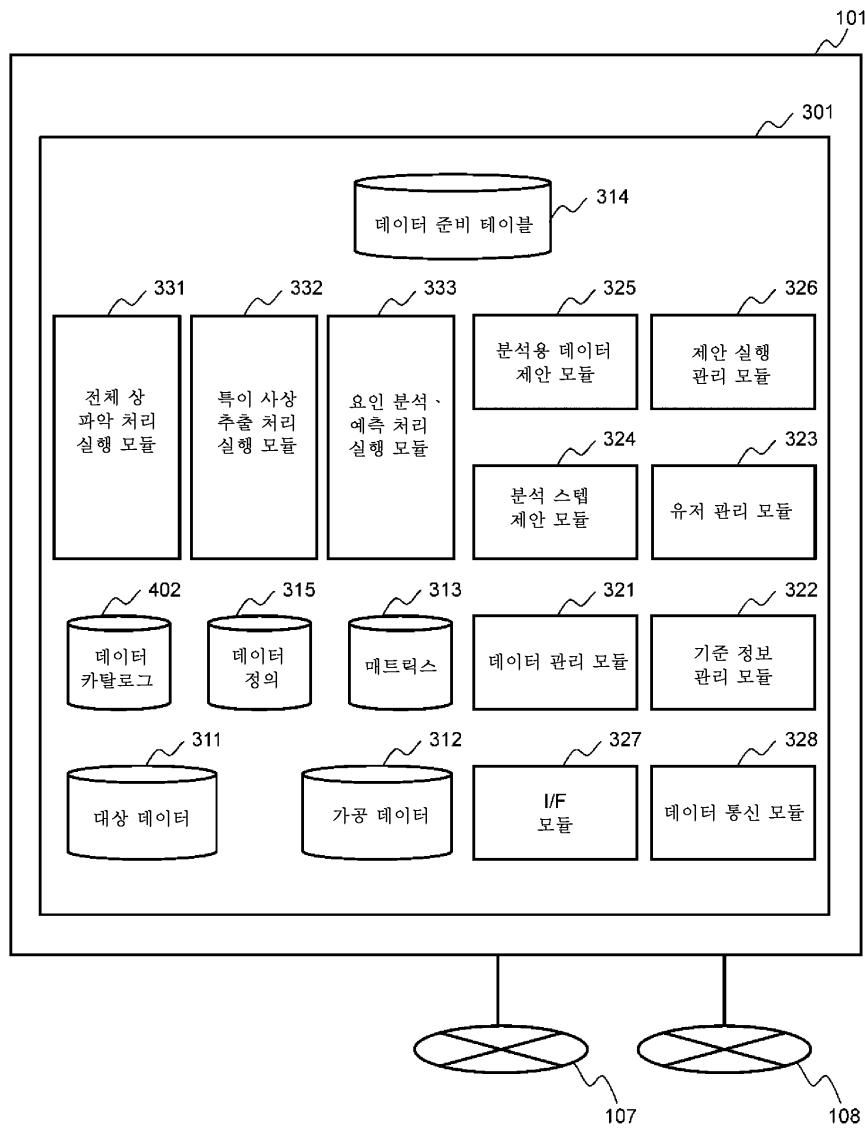
도면1



도면2



도면3



도면4

401

【분석 목적 정보】

■ 분석 스텝(필수)

- 이하 중 어느 것을 지정
(1) 전체 상 파악 (2) 특이 사상 추출 (3) 요인 분석 · 예측

■ KPI (필수)

- 목적에 대해 주목하는 KPI의 데이터 명칭

□ 착안 데이터 항목(임의)

- KPI 데이터를 평가하기 위한 축으로서 착안하는 데이터의 명칭
- KPI 데이터와 동시에 감시하고자 하는 데이터의 명칭

402

【데이터 카탈로그】

보유 데이터

(데이터 칼럼명), (명칭·의미), (보충 정보)

arrival time, 도착 시각, ...

KRT, 킬로정, ...

가공 데이터

(데이터명), (산출식), (명칭·의미), (보충 정보)

delay, “실적 시각”-“계획 시각”, 지역 시분, ...

403

【4W 사전】

(키워드), (카테고리)

시각, when

date, when

킬로정, where

Kilometrage, where

역 코드, where

404

【데이터 추이 패턴】

(데이터명·종별), (카테고리), (데이터형), (값 변화 폭), (하한), (상한)

킬로정, where, integer, 1, 0, 300

도착 시각(분), when, integer, 5, 0, 1440

도면5a

| 순위 | 식별 정보 | KPI | 작업 데이터 1 (When) | 작업 데이터 2 (Where) | 작업 데이터 3 (What) | 작업 데이터 4 (Who) | 레코드수 | 출력 파일 |
|-----|----------|-----|-----------------------|------------------------|-----------------------|----------------------|------|----------|
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |

501

도면5b

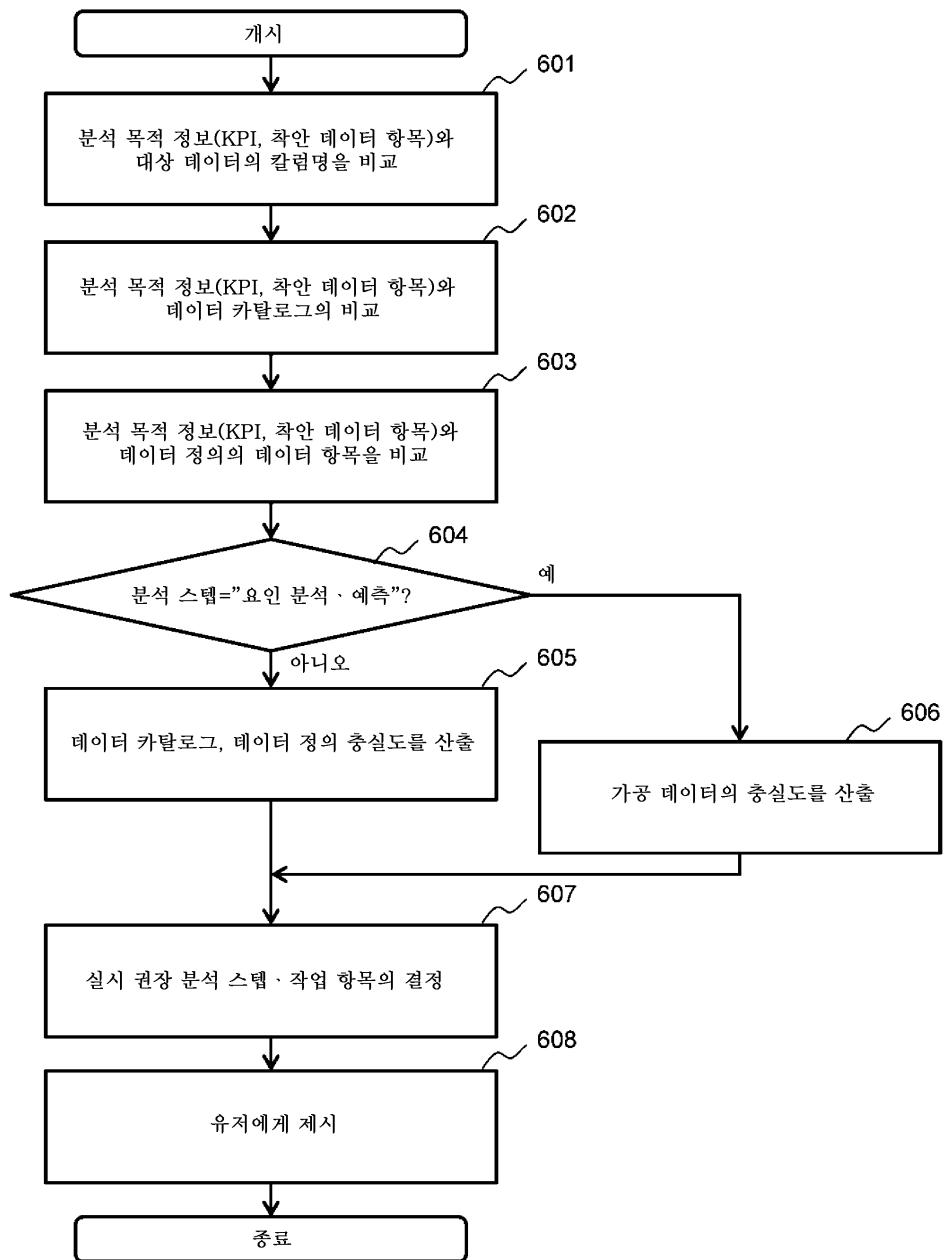
도면5c

| 순위 | 식별 정보 | 목적 변수 (KPI) | 설명 변수 | 레코드수 | 유용도1 | 유용도2 | 유용도3 | 유용도4 | 출력 파일 |
|-----|----------|-------------------|----------|------|------|------|------|------|----------|
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

도면5d

| 식별 정보 | KPI | 데이터 조합 | 이용 횟수 | 이용 인원수 | 갱신 일시 |
|-------|-----|--------|-------|--------|-------|
| ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... |

도면6



도면7a

| 701 | 분석 목적 정보와의 일치도 | | |
|-----------------------------|----------------|---------|---------------------------------------|
| | 높음 | 낮음 | |
| 데이터 카탈로그/ 데이터 정의의 충실도 | 높음 | 전체 상 파악 | 데이터 추가 |
| | 낮음 | 전체 상 파악 | 데이터의 이해 촉진, 데이터 카탈로그/ 데이터 정의 확충 |

도면7b

702

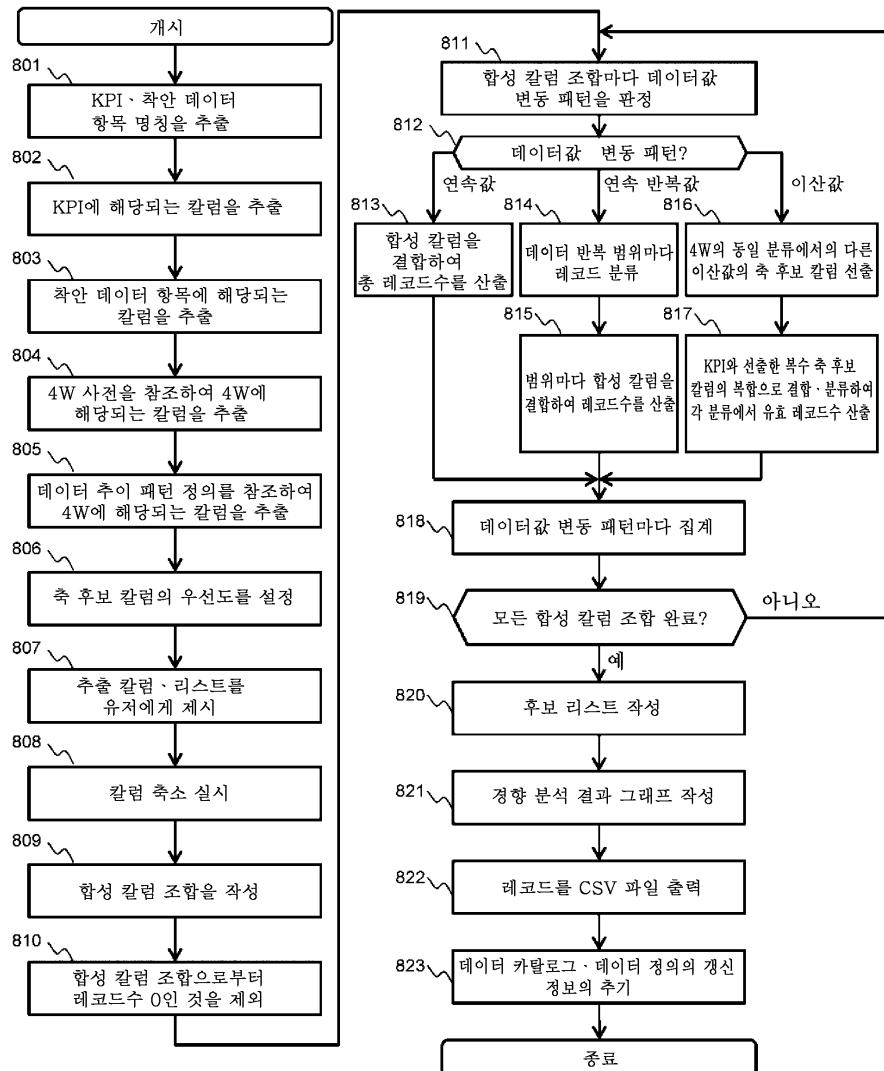
| | | 분석 목적 정보와의 일치도 | |
|-------------------------------|----|----------------|---------|
| | | 높음 | 낮음 |
| 분석 목적 정보에 일치하는 데이터의 충실도 | 높음 | 특이 사상 추출 | 전체 상 파악 |
| | 낮음 | 데이터 추가 요구 | 전체 상 파악 |

도면7c

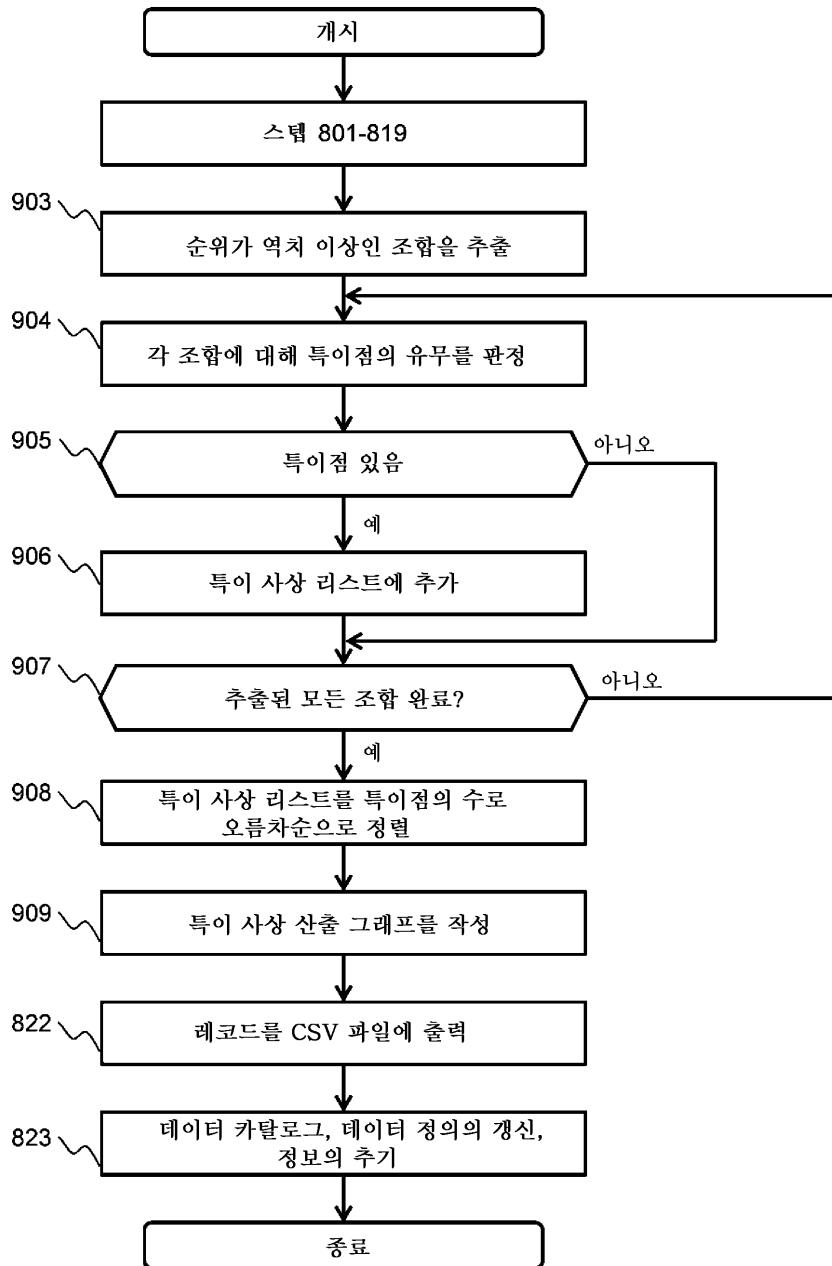
703

| | | 분석 목적 정보와의 일치도 | |
|----------------|----|----------------|-----------------------|
| | | 높음 | 낮음 |
| 가공 데이터의 충실도 | 높음 | 요인 분석 · 예측 | 전체 상 파악 |
| | 낮음 | 가공 데이터 확충 | 전체 상 파악, 가공 데이터 확충 |

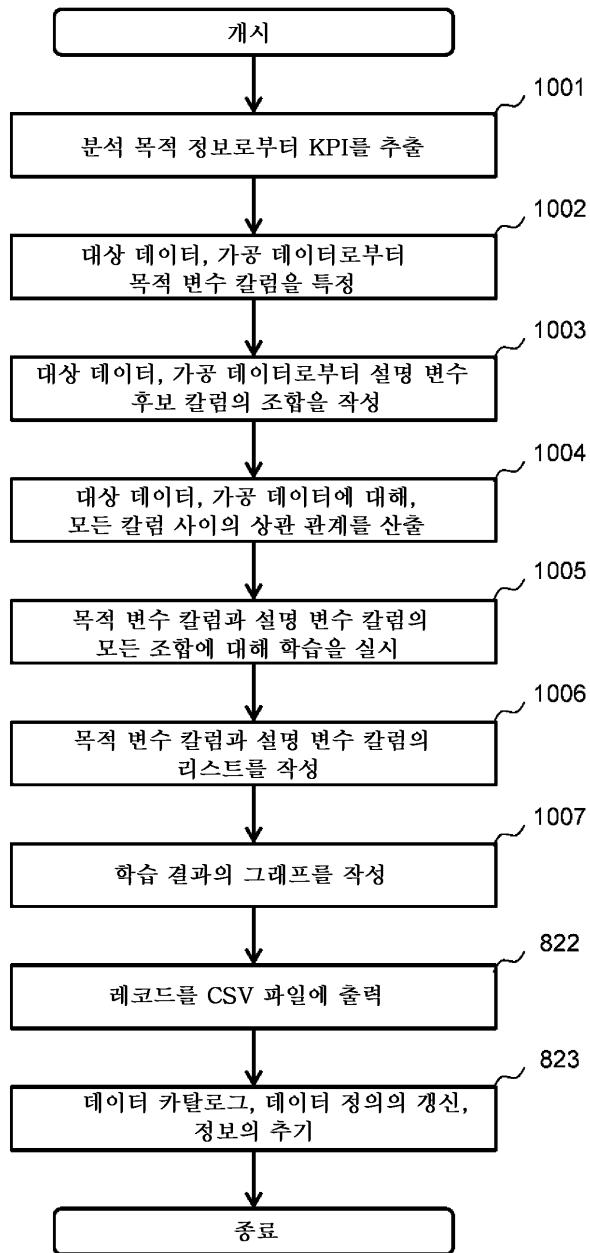
도면8



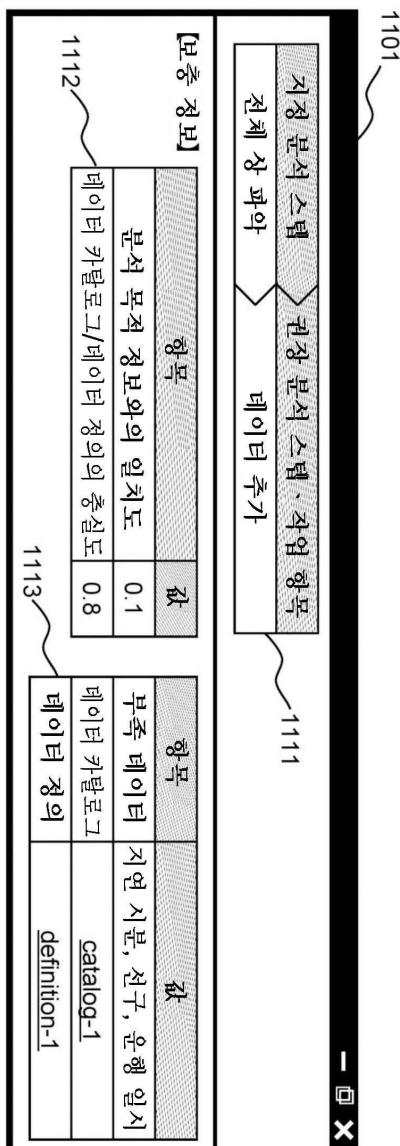
도면9



도면10



도면 11a



도면11b

1102 ~ 1121

— ☐ X —

| 분석 스텝 | | 전체 상_파악 | | | |
|-------|-------|--------------|-------|------|-------|
| # | KPI | When | Where | What | Who |
| 1 | 지연 지분 | •도착 시각 •역 | •천구 | — | 1,500 |
| 2 | 지연 지분 | •발차 시각 | •킬로정 | — | 1,200 |
| 3 | 지연 지분 | •운행 일시 | — | •일번 | 900 |
| ... | ... | ... | ... | ... | ... |

1122

【참고 그레프】 #1 #2 #3

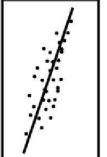
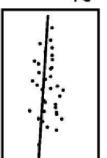
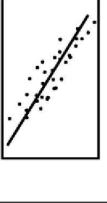
도면11c

1103

- ◉ X

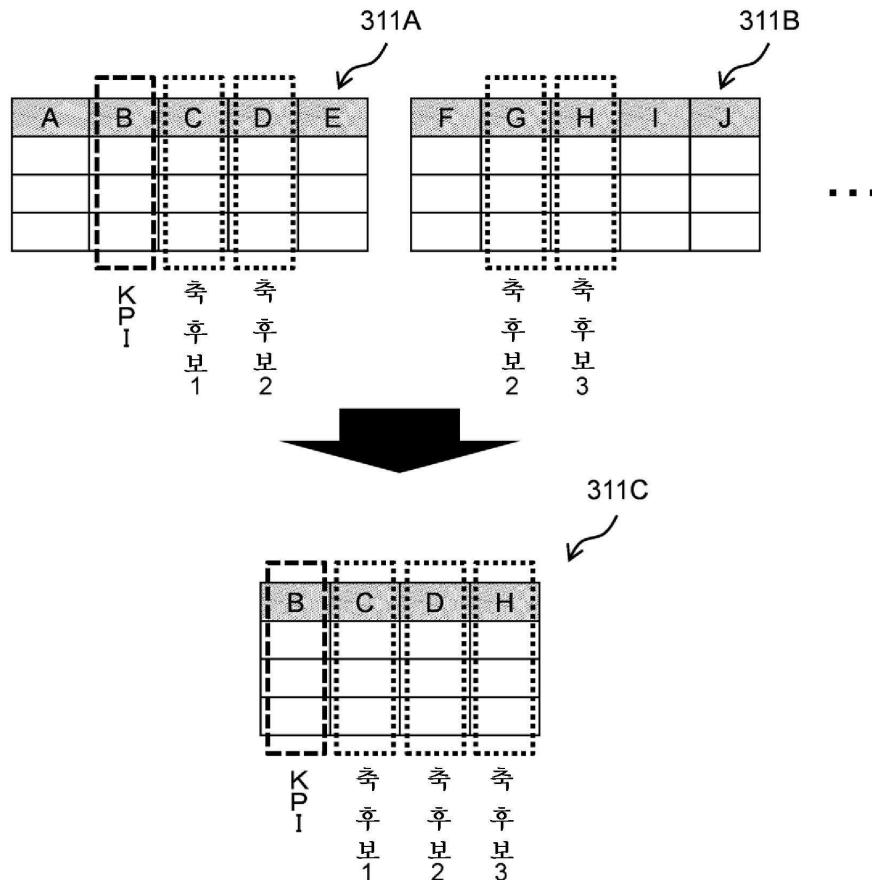
| 분석 스텝 | | 요인 분석 · 예측 | | | | | |
|-------|----------------|------------|--------|--------|-----|-------------|-----------|
| # | 목적 변수 (KPI) | 설명 변수1 | 설명 변수2 | 설명 변수3 | ... | 출처 파일 횟수 | 이용 인원수 |
| 1 | 지연 시간 | 승차율 | 역 혼잡도 | 운행 횟수 | | csvfile-1 | 20 |
| 2 | 지연 시간 | 기온 | 습도 | 강수량 | | csvfile-2 | 10 |
| 3 | 지연 시간 | 시간대 | — | — | | csvfile-3 | 15 |
| ... | ... | ... | ... | ... | ... | ... | 5 |

1131

【참고 그레프】 #1  #2  #3 

1132

도면12



도면13

