



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0012275
(43) 공개일자 2023년01월26일

(51) 국제특허분류(Int. Cl.)

G06F 16/2453 (2019.01) G06F 16/2455

(2019.01)

G06F 16/2458 (2019.01) G06F 16/28 (2019.01)

G06N 20/00 (2019.01)

(52) CPC특허분류

G06F 16/24545 (2019.01)

G06F 16/24537 (2019.01)

(21) 출원번호 10-2021-0092850

(22) 출원일자 2021년07월15일

심사청구일자 2021년07월15일

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

정연돈

서울특별시 중구 청계천로 400, 105-812

이협진

서울특별시 성동구 왕십리로 410, 112동 601호

(74) 대리인

김홍석

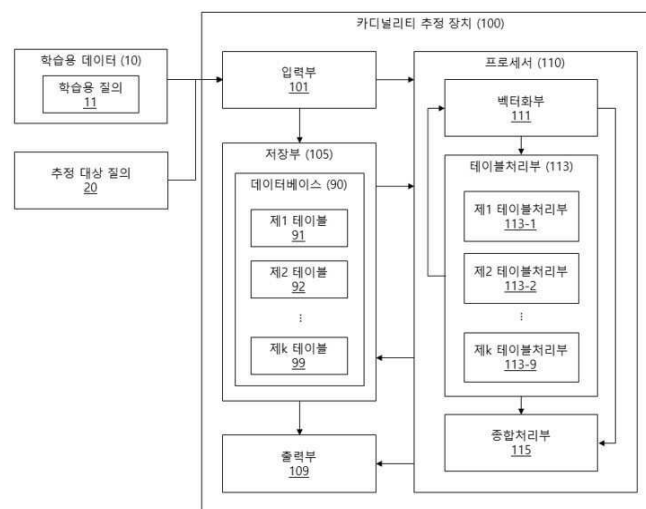
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 조인 질의를 위한 카디널리티 추정 장치 및 방법

(57) 요약

조인 질의를 위한 카디널리티 추정 장치 및 방법에 관한 것으로, 조인 질의를 위한 카디널리티 추정 장치는, 다수의 테이블에 대한 조인 질의를 포함하는 추정 대상 질의를 입력 받는 입력부 및 상기 추정 대상 질의에 대한 제1 질의 벡터를 획득하고, 상기 제1 질의 벡터를 이용하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하고, 상기 각각의 테이블에 대한 카디널리티 추정 값을 기반으로 상기 추정 대상 질의에 대한 제2 질의 벡터를 획득하고, 상기 제2 질의 벡터를 이용하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 프로세서를 포함할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06F 16/2456 (2019.01)

G06F 16/2462 (2019.01)

G06F 16/284 (2019.01)

G06N 20/00 (2021.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711127329
과제번호	2020R1A2C2013286
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	대규모 분산 딥러닝을 위한 인메모리 텐서 데이터베이스
기 여 율	1/2
과제수행기관명	고려대학교 산학협력단
연구기간	2021.03.01 ~ 2022.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711126325
과제번호	2020-0-01819-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	ICT명품인재양성(고려대학교)
기 여 율	1/2
과제수행기관명	고려대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

다수의 테이블에 대한 조인 질의를 포함하는 추정 대상 질의를 입력 받는 입력부; 및

상기 추정 대상 질의에 대한 제1 질의 벡터를 획득하고, 상기 제1 질의 벡터를 이용하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하고, 상기 각각의 테이블에 대한 카디널리티 추정 값을 기반으로 상기 추정 대상 질의에 대한 제2 질의 벡터를 획득하고, 상기 제2 질의 벡터를 이용하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 프로세서;를 포함하는 카디널리티 추정 장치.

청구항 2

제1항에 있어서,

상기 제1 질의 벡터는, 상기 추정 대상 질의에서 이용된 연산자에 대한 정보와, 상기 이용된 연산자를 통한 연산에 이용된 값을 포함하는 카디널리티 추정 장치.

청구항 3

제1항에 있어서,

상기 제2 질의 벡터는, 복수의 테이블에 대한 복수의 영역을 포함하고, 상기 복수의 영역은 각각 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값과, 상기 추정 대상 질의에서 이용된 연산자에 대한 정보와, 상기 이용된 연산자를 통한 연산에 이용된 값을 포함하는 카디널리티 추정 장치.

청구항 4

제1항에 있어서,

상기 프로세서는, 상기 제1 질의 벡터를 상기 추정 대상 질의의 대상이 되는 테이블에 대응하는 제1 학습 모델에 입력하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하거나, 또는 상기 제2 질의 벡터를 조인 질의에 관한 카디널리티 추정을 위한 제2 학습 모델에 입력하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 카디널리티 추정 장치.

청구항 5

제4항에 있어서,

상기 입력부는, 학습용 질의를 포함하는 학습용 데이터를 입력 받고,

상기 프로세서는, 상기 학습용 질의를 이용하여 상기 제1 학습 모델 또는 상기 제2 학습 모델을 학습하는 카디널리티 추정 장치.

청구항 6

다수의 테이블에 대한 조인 질의를 포함하는 추정 대상 질의를 입력 받는 단계;

상기 추정 대상 질의에 대한 제1 질의 벡터를 획득하는 단계;

상기 제1 질의 벡터를 이용하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하는 단계;

상기 각각의 테이블에 대한 카디널리티 추정 값을 기반으로 상기 추정 대상 질의에 대한 제2 질의 벡터를 획득하는 단계; 및

상기 제2 질의 벡터를 이용하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 단계;를 포함하는 카디널리티 추정 방법.

청구항 7

제6항에 있어서,

상기 제1 질의 벡터는, 상기 추정 대상 질의에서 이용된 연산자에 대한 정보와, 상기 이용된 연산자를 통한 연산에 이용된 값을 포함하는 카디널리티 추정 방법.

청구항 8

제6항에 있어서,

상기 제2 질의 벡터는, 복수의 테이블에 대한 복수의 영역을 포함하고, 상기 복수의 영역은 각각 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값과, 상기 추정 대상 질의에서 이용된 연산자에 대한 정보와, 상기 이용된 연산자를 통한 연산에 이용된 값을 포함하는 카디널리티 추정 방법.

청구항 9

제6항에 있어서,

상기 제1 질의 벡터를 이용하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하는 단계는,

상기 제1 질의 벡터를 상기 추정 대상 질의의 대상이 되는 테이블에 대응하는 제1 학습 모델에 입력하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하는 단계;를 포함하거나, 또는

상기 제2 질의 벡터를 이용하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 단계는,

상기 제2 질의 벡터를 상기 조인 질의에 관한 카디널리티의 추정을 위한 제2 학습 모델에 입력하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 단계;를 포함하는 카디널리티 추정 방법.

청구항 10

제9항에 있어서,

학습용 질의를 포함하는 학습용 데이터를 입력 받는 단계; 및

상기 제1 학습 모델 또는 상기 제2 학습 모델을 상기 학습용 질의를 이용하여 훈련시키는 단계;를 더 포함하는 카디널리티 추정 방법.

청구항 11

다수의 테이블에 대한 조인 질의를 포함하는 학습용 질의를 입력 받는 입력부; 및

상기 학습용 질의에 대한 제1 질의 벡터를 획득하고, 상기 제1 질의 벡터를 이용하여 각각의 테이블에 대한 카디널리티 추정 값의 획득을 위한 제1 학습 모델을 훈련시키고, 상기 각각의 테이블에 대한 실제 카디널리티 값 또는 추정 카디널리티 값을 이용하여 상기 학습용 질의에 대한 제2 질의 벡터를 획득하고, 상기 제2 질의 벡터를 이용하여 조인 질의에 관한 카디널리티 추정을 위한 제2 학습 모델을 훈련시키는 프로세서;를 포함하는 카디널리티 추정 장치.

청구항 12

다수의 테이블에 대한 조인 질의를 포함하는 학습용 질의를 입력 받는 단계;

상기 학습용 질의에 대한 제1 질의 벡터를 획득하는 단계;

상기 제1 질의 벡터를 이용하여, 상기 학습용 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값의 획득을 위한 제1 학습 모델을 훈련시키는 단계;

상기 각각의 테이블에 대한 카디널리티 추정 값을 기반으로 상기 학습용 질의에 대한 제2 질의 벡터를 획득하는 단계; 및

상기 제2 질의 벡터를 이용하여 조인 질의에 관한 카디널리티 추정을 위한 제2 학습 모델을 훈련시키는 단계;를 포함하는 카디널리티 추정 방법.

발명의 설명

기술 분야

[0001] 조인 질의를 위한 카디널리티 추정 장치 및 방법에 관한 것이다.

배경 기술

[0002] 관계형 데이터베이스(Relational Database)는 관계형 모델을 기반으로 테이블을 이용하여 데이터를 표현하는 데이터베이스의 일 유형으로, 검색의 편의성 등의 장점이 있어 현재 가장 널리 사용되는 데이터베이스이다. 관계형 데이터베이스의 테이블은 열과 행으로 이루어지는데, 각각의 열에는 특정 속성(attribute)에 해당하는 값인 속성 값이 마련되고, 각각의 행에는 서로 관계된 속성 값의 집합인 튜플(tuple)이 마련된다. 통상 이러한 속성의 개수를 차수(degree)라고 하고, 튜플의 개수를 카디널리티(Cardinality)라고 한다. 사용자 등으로부터 질의(쿼리, query)가 입력되면, 데이터베이스 관리 시스템(DBMS: Database Management System) 등은 질의에 대응하는 응답을 검색하여 획득하는데, 이때 응답의 검색을 최적화하기 위해 적절한 쿼리 계획을 생성한다. 이 경우, 시스템은 처리될 총 행(튜플)의 개수, 즉 카디널리티를 추정하고, 추정된 카디널리티를 비용 모델의 입력 값으로 이용하여 최적의 쿼리 계획을 결정한다. 따라서, 적절한 카디널리티의 추정은 관계형 데이터베이스 검색의 최적화에 지대한 영향을 미친다. 종래의 데이터베이스 관리 시스템은 이러한 카디널리티의 추정을 위해 데이터(들)가 서로 독립적이고 균등 분포를 갖고 있다고 가정하였다. 그러나, 실제 데이터들은 독립적이지 않을 수도 있고 또한 균등하게 분포되어 있지 않을 수도 있으므로, 종래의 데이터베이스 관리 시스템은 카디널리티를 정확하게 추정하기 어려웠으며, 이는 데이터베이스 내의 각각의 자료들에 대한 신속한 검색이나 최적의 관리에 장애가 되어 왔다. 한편, 사용자는 서로 상이한 둘 이상의 테이블에 대한 처리를 요구하는 질의(조인 질의)를 입력할 수도 있다. 이 경우, 데이터베이스 관리 시스템은 둘 이상의 테이블 간에 연관된 속성(일례로 공통 속성) 등을 이용하여 양 테이블을 결합하여 데이터를 검색하였다. 그런데, 이러한 조인 질의의 경우에도, 질의에 대한 응답 확보의 최적화를 위해 적절한 카디널리티가 추정되어야 하나, 종래의 시스템은 이러한 조인 질의의 처리가 불가능하거나, 조인 질의의 처리가 가능하다고 하더라도 조인 질의에 대한 카디널리티 값을 추정하는데 긴 처리 시간을 요구하여 입력된 질의에 대응하는 응답을 신속하게 획득할 수 없는 문제점이 존재하였다.

발명의 내용

해결하려는 과제

[0003] 조인 질의의 처리를 위한 카디널리티를 신속하면서도 높은 정확도로 추정할 수 있는 카디널리티 추정 장치 및 방법을 제공하는 것을 해결하고자 하는 과제로 한다.

[0004] 또한, 본 발명은 한국연구재단의 '프라이버시 침해 없이 감염병을 예방하는 데이터 기반 시스템' 과제(과제번호: 2020K2A9A1A0109589411), 한국연구재단의 '대규모 분산 딥러닝을 위한 인메모리 텐서 데이터베이스' 과제(과제번호: 2020R1A2C201328612), 및 정보통신기획평가원의 'ICT명품인재양성' 과제(과제번호: 20200018190021001)의 수행 결과물이다.

과제의 해결 수단

[0005] 상술한 과제를 해결하기 위하여 카디널리티 추정 장치 및 방법이 제공된다.

[0006] 카디널리티 추정 장치는, 다수의 테이블에 대한 조인 질의를 포함하는 추정 대상 질의를 입력 받는 입력부 및 상기 추정 대상 질의에 대한 제1 질의 벡터를 획득하고, 상기 제1 질의 벡터를 이용하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하고, 상기 각각의 테이블에 대한 카디널리티 추정 값을 기반으로 상기 추정 대상 질의에 대한 제2 질의 벡터를 획득하고, 상기 제2 질의 벡터를 이용하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 프로세서를 포함할 수 있다.

[0007] 상기 제1 질의 벡터는, 상기 추정 대상 질의에서 이용된 연산자에 대한 정보와, 상기 이용된 연산자를 통한 연산에 이용된 값을 포함할 수도 있다.

[0008] 상기 제2 질의 벡터는, 복수의 테이블에 대한 복수의 영역을 포함하고, 상기 복수의 영역은 각각 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값과, 상기 추정 대상 질의에서 이용된 연산자에 대한 정보와, 상기 이용된 연산자를 통한 연산에 이용된 값을 포함할 수도 있다.

- [0009] 상기 프로세서는, 상기 제1 질의 벡터를 상기 추정 대상 질의의 대상이 되는 테이블에 대응하는 제1 학습 모델에 입력하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하거나, 또는 상기 제2 질의 벡터를 상기 조인 질의에 관한 카디널리티 추정을 위한 제2 학습 모델에 입력하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득할 수도 있다.
- [0010] 상기 입력부는, 학습용 질의를 포함하는 학습용 데이터를 입력 받고, 상기 프로세서는, 상기 학습용 질의를 이용하여 상기 제1 학습 모델 또는 상기 제2 학습 모델을 학습할 수도 있다.
- [0011] 카디널리티 추정 방법은, 다수의 테이블에 대한 조인 질의를 포함하는 추정 대상 질의를 입력 받는 단계, 상기 추정 대상 질의에 대한 제1 질의 벡터를 획득하는 단계, 상기 제1 질의 벡터를 이용하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하는 단계, 상기 각각의 테이블에 대한 카디널리티 추정 값을 기반으로 상기 추정 대상 질의에 대한 제2 질의 벡터를 획득하는 단계 및 상기 제2 질의 벡터를 이용하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 단계를 포함할 수 있다.
- [0012] 상기 제1 질의 벡터는, 상기 추정 대상 질의에서 이용된 연산자에 대한 정보와, 상기 이용된 연산자를 통한 연산에 이용된 값을 포함하는 것도 가능하다.
- [0013] 상기 제2 질의 벡터는, 복수의 테이블에 대한 복수의 영역을 포함하고, 상기 복수의 영역은 각각 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값과, 상기 추정 대상 질의에서 이용된 연산자에 대한 정보와, 상기 이용된 연산자를 통한 연산에 이용된 값을 포함할 수도 있다.
- [0014] 상기 제1 질의 벡터를 이용하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하는 단계는, 상기 제1 질의 벡터를 상기 추정 대상 질의의 대상이 되는 테이블에 대응하는 제1 학습 모델에 입력하여 상기 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값을 획득하는 단계를 포함할 수 있다.
- [0015] 또한, 상기 제2 질의 벡터를 이용하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 단계는, 상기 제2 질의 벡터를 상기 조인 질의에 관한 카디널리티의 추정을 위한 제2 학습 모델에 입력하여 상기 조인 질의에 대한 카디널리티 추정 값을 획득하는 단계를 포함할 수도 있다.
- [0016] 카디널리티 추정 방법은, 학습용 질의를 포함하는 학습용 데이터를 입력 받는 단계 및 상기 제1 학습 모델 또는 상기 제2 학습 모델을 상기 학습용 질의를 이용하여 훈련시키는 단계를 더 포함할 수도 있다.
- [0017] 카디널리티 추정 장치는, 다수의 테이블에 대한 조인 질의를 포함하는 학습용 질의를 입력 받는 입력부 및 상기 학습용 질의에 대한 제1 질의 벡터를 획득하고, 상기 제1 질의 벡터를 이용하여 각각의 테이블에 대한 카디널리티 추정 값의 획득을 위한 제1 학습 모델을 훈련시키고, 상기 각각의 테이블에 대한 실제 카디널리티 값 또는 추정 카디널리티 값을 이용하여 상기 학습용 질의에 대한 제2 질의 벡터를 획득하고, 상기 제2 질의 벡터를 이용하여 조인 질의에 관한 카디널리티 추정을 위한 제2 학습 모델을 훈련시키는 프로세서를 포함하는 것도 가능하다.
- [0018] 카디널리티 추정 방법은, 다수의 테이블에 대한 조인 질의를 포함하는 학습용 질의를 입력 받는 단계, 상기 학습용 질의에 대한 제1 질의 벡터를 획득하는 단계, 상기 제1 질의 벡터를 이용하여, 상기 학습용 질의의 대상이 되는 각각의 테이블에 대한 카디널리티 추정 값의 획득을 위한 제1 학습 모델을 훈련시키는 단계, 상기 각각의 테이블에 대한 카디널리티 추정 값을 기반으로 상기 학습용 질의에 대한 제2 질의 벡터를 획득하는 단계 및 상기 제2 질의 벡터를 이용하여 조인 질의에 관한 카디널리티 추정을 위한 제2 학습 모델을 훈련시키는 단계를 포함할 수도 있다.
- 발명의 효과**
- [0019] 상술한 카디널리티 추정 장치 및 방법에 의하면, 보다 짧은 시간에 높은 정확도로 조인 질의를 위한 카디널리티를 추정할 수 있게 되는 효과를 얻을 수 있다.
- [0020] 상술한 카디널리티 추정 장치 및 방법에 의하면, 카디널리티의 보다 신속하고 정확한 추정이 가능하므로 입력된 질의에 대한 최적의 실행 계획을 획득할 수 있게 되고, 이에 따라 관계형 데이터베이스의 질의 처리 및 검색의 성능이 향상되고 효율성이 개선되는 효과를 얻을 수 있다.
- [0021] 상술한 카디널리티 추정 장치 및 방법에 의하면, 상대적으로 간단하고 가벼운 모델만으로도 조인 질의를 위한 카디널리티 값을 보다 정확하게 결정할 수 있게 되므로, 카디널리티 추정 시간이 단축되고 아울러 하드웨어 리

소스 등의 사용량을 절감할 수 있는 경제적 이점도 존재한다.

도면의 간단한 설명

[0022] 도 1은 카디널리티 추정 장치의 블록도이다.

도 2는 각각의 테이블에 대응하는 학습 모델의 처리를 위한 조인 질의의 벡터화의 일례를 설명하기 위한 도면이다.

도 3은 다수의 테이블 간의 관계에 대한 학습 모델의 처리를 위한 조인 질의의 벡터화의 일례를 설명하기 위한 도면이다.

도 4는 프로세서의 학습 동작의 일례를 설명하기 위한 도면이다.

도 5는 카디널리티 추정 방법에 대한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0023] 이하 명세서 전체에서 동일 참조 부호는 특별한 사정이 없는 한 동일 구성요소를 지칭한다. 이하에서 사용되는 '부'가 부가된 용어는, 소프트웨어 및/또는 하드웨어로 구현될 수 있으며, 실시예에 따라 하나의 '부'가 하나의 물리적 또는 논리적 부품으로 구현되거나, 복수의 '부'가 하나의 물리적 또는 논리적 부품으로 구현되거나, 하나의 '부'가 복수의 물리적 또는 논리적 부품들로 구현되는 것도 가능하다. 명세서 전체에서 어떤 부분이 다른 부분과 연결되어 있다고 할 때, 이는 어떤 부분과 다른 부분이 상호 간에 물리적으로 연결되었음을 의미할 수도 있고, 및/또는 전기적으로 연결되었음을 의미할 수도 있다. 또한, 어떤 부분이 다른 부분을 포함한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 부분 이외의 또 다른 부분을 제외하는 것이 아니며, 설계자의 선택에 따라서 또 다른 부분을 더 포함할 수 있음을 의미한다. 제1 내지 제N(N은 1 이상의 자연수) 등의 표현은, 적어도 하나의 부분(들)을 다른 부분(들)으로부터 구분하기 위한 것으로, 특별한 기재가 없는 이상 이들이 순차적임을 반드시 의미하지는 않는다. 또한 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함할 수 있다.

[0024] 이하 도 1 내지 도 4를 참조하여 조인 질의를 위한 카디널리티 추정 장치의 여러 실시예에 대해서 설명하도록 한다.

[0025] 도 1은 카디널리티 추정 장치의 블록도이다.

[0026] 도 1에 도시된 바를 참조하면, 카디널리티 추정 장치(100)는 일 실시예에 있어서 입력부(101), 저장부(105), 출력부(109) 및 프로세서(110)를 포함할 수 있다. 입력부(101), 저장부(105), 출력부(109) 및 프로세서(110) 중 적어도 둘은 회로선, 케이블 또는 무선 네트워크 등을 통해 일방으로 또는 쌍방으로 데이터나 명령/지시 등을 전달할 수 있도록 마련된다. 필요에 따라서, 입력부(101), 저장부(105) 및 출력부(109) 중 적어도 하나는 생략 가능하다.

[0027] 입력부(101)는, 학습용 데이터(10) 및 추정 대상 질의(20) 중 적어도 하나를, 사용자의 수동 조작에 따라 사용자로부터 직접 수신하거나 및/또는 다른 장치(일례로 휴대용 메모리 장치나 다른 정보처리장치 등)로부터 전달받아 획득할 수 있다. 입력부(101)는, 실시예에 따라 카디널리티 추정 장치(100)와 일체형으로 마련된 것일 수도 있고, 또는 물리적으로 분리 가능하게 마련된 것일 수도 있다. 입력부(101)는, 예를 들어, 키보드, 마우스, 태블릿, 터치 스크린, 터치 패드, 트랙 볼, 트랙패드, 스캐너 장치, 영상 촬영 모듈, 초음파 스캐너, 동작 감지 센서, 진동 센서, 수광 센서, 감압 센서, 근접 센서 및/또는 마이크로 폰 등을 포함할 수 있다. 또한 실시예에 따라, 입력부(101)는, 외부의 다른 장치(일례로 휴대용 메모리 장치 등)로부터 데이터 등의 수신에 가능한 데이터 입출력 단자나, 외부의 다른 장치와 유무선 통신 네트워크를 통해 연결되는 통신 모듈(일례로 랜 카드, 근거리 통신 모듈 또는 이동통신 모듈 등) 등을 포함할 수도 있다.

[0028] 학습용 데이터(10)는 프로세서(110)에 의한 학습 모델(일례로 도 4의 113-1a, 113-2a, 113-3a, 115a 등)의 훈련을 위해 이용되는 데이터로, 구체적으로 카디널리티 추정을 위한 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)의 훈련에 이용될 수 있다. 학습용 데이터(10)는, 데이터베이스(90) 내의 데이터의 관리, 검색, 확인 또는 조작 등을 위한 질의(11, query)를 하나 이상 포함할 수 있다. 여기서, 하나 이상의 학습용 질의(11)는 오직 하나의 테이블(릴레이션으로 표현될 수도 있음. 91 내지 99 중 하나)에 대한 질의만을 포함할 수도 있고, 또는 둘 이상의 테이블(91 내지 99 중 둘 이상)에 대한 질의(이하 조인 질의)를 포함할 수도 있다. 이하 각각의 테이블(91 내지 99)을 제1 내지 제k 테이블(k는 1 이상의 자연수)로 칭한다. 학습용 데이터(10)는, 실시예에 따라, 입

력된 학습용 질의(11) 및 학습용 질의(11)의 대상이 되는 테이블(91 내지 99)을 통해 획득 가능한 실제 카디널리티 값(미도시)을 더 포함할 수도 있다. 실제 카디널리티 값은 제1 내지 제k 테이블(91 내지 99) 중 적어도 하나에 대응하는 값을 포함할 수도 있고, 및/또는 조인 질의에 의해 처리될 둘 이상의 테이블(91 내지 99 중 적어도 둘)에 대응하는 값을 포함할 수도 있다. 이들 실제 카디널리티 값은 테이블처리부(113) 또는 종합처리부(115)의 지도 학습을 위해 이용될 수 있다. 예를 들어, 실제 카디널리티 값은, 제1 내지 제k 테이블(91 내지 99) 각각에 대응하는 제1 내지 제k 테이블처리부(113-1 내지 113-9)의 학습 과정에서 해당 학습용 질의(11)에 대응하는 출력 값으로 이용될 수도 있고, 종합처리부(115)의 조인 질의에 관한 학습 과정에서 해당 조인 질의에 대응하는 출력 값으로 이용될 수도 있다.

[0029] 추정 대상 질의(20)는 카디널리티 추정 장치(100)의 사용자나 설계자 등이 추정 대상 질의(20)의 목적이 되는 적어도 하나의 테이블(91 내지 99)의 관리나 검색 등을 위해 입력하는 질의로, 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)에 의한 카디널리티 값의 추정에 이용될 수 있다. 추정 대상 질의(20)도 오직 하나의 테이블(91 내지 99 중 하나)에 대한 질의만을 포함할 수도 있고, 및/또는 둘 이상의 테이블(91 내지 99 중 둘 이상)에 대한 조인 질의를 포함할 수도 있다. 추정 대상 질의(20)가 오직 하나의 테이블(91 내지 99 중 하나)에 대한 질의만을 포함하는 경우에는 테이블처리부(113)에 의해 해당 추정 대상 질의(20) 및 해당 테이블(91 내지 99 중 하나)에 대한 카디널리티의 추정 값(도 3의 91b, 92b)이 획득되고, 추정 대상 질의(20)가 조인 질의를 포함하는 경우에는 테이블처리부(113) 및 종합 처리부(115)에 의해 해당 추정 대상 질의(20) 및 해당 둘 이상의 테이블(91 내지 99 중 둘 이상)에 대한 카디널리티의 추정 값(도 4의 99)이 획득될 수 있다.

[0030] 실시예에 따라서, 학습용 데이터(10) 및 추정 대상 질의(20)는 상황에 따라 동시에 입력될 수도 있고 순차적으로 입력될 수도 있다. 또한, 필요에 따라, 카디널리티 추정 장치(100)는 학습용 데이터(10)만을 입력 받을 수도 있고, 추정 대상 질의(20)만을 입력 받을 수도 있다. 학습용 데이터(10)만을 입력 받는 경우, 카디널리티 추정 장치(100)의 프로세서(110)는 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)을 획득한 후, 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)을 출력부(109) 등을 통해 메모리 장치나 다른 정보처리장치(미도시)로 전달하고, 다른 정보처리장치 등은 추정 대상 질의(20)를 입력 받고 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)을 이용하여 추정 대상 질의(20)에 대응하는 카디널리티를 추정할 수도 있다. 추정 대상 질의(20)만을 입력 받는 경우, 카디널리티 추정 장치(100)는 입력부(101) 등을 통해 다른 정보처리장치에 의해 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)을 획득하고, 프로세서(110)는 이를 이용하여 추정 대상 질의(20)에 대응하는 카디널리티를 추정할 수도 있다.

[0031] 저장부(105)는 카디널리티 추정 장치(100)의 동작에 필요한 데이터, 명령 또는 프로그램(소프트웨어, 애플리케이션 또는 앱 등으로 표현될 수도 있음)을 일시적 또는 비일시적으로 저장할 수 있다. 예를 들어, 저장부(105)는 학습용 데이터(10)의 학습용 질의(11) 또는 추정 대상 질의(20)를 일시적 또는 비일시적으로 저장할 수도 있고, 학습용 질의(11) 또는 추정 대상 질의(20)의 대상이 되는 데이터베이스(90)를 저장할 수도 있다. 여기서, 데이터베이스(90)는 관계형 데이터베이스일 수 있으며, 하나 또는 둘 이상의 테이블(91 내지 99)을 포함할 수 있다. 둘 이상의 테이블(91 내지 99)이 마련된 경우, 적어도 하나의 테이블(91 내지 99)의 속성의 종류는 다른 적어도 하나의 테이블(91 내지 99)의 속성과 모두 동일할 수도 있고, 일부는 동일하고 일부는 상이할 수도 있으며, 또는 모두 상이할 수도 있다. 또한, 각각의 테이블(91 내지 99)의 속성의 개수(차수) 역시 모두 동일할 수도 있고, 모두 상이할 수도 있으며, 일부의 테이블(91 내지 99) 간에만 동일할 수도 있다. 또한 복수의 테이블(91 내지 99)의 전부 또는 일부는 상황에 따라 동일한 데이터를 공통적으로 가지고 있을 수도 있다. 저장부(105)는 프로세서(110)의 동작에 필요한 적어도 하나의 프로그램이나 프로세서(110)에 의해 훈련될 또는 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등) 등을 저장할 수도 있다. 여기서, 프로세서(110)의 동작에 필요한 프로그램은, 설계자에 의해 직접 작성되어 저장부(105)에 저장 또는 수정된 것일 수도 있고, 휴대용 메모리 장치나 디브이디 등의 기록 매체를 이용하여 입력부(101)를 통해 입력되어 저장 또는 갱신된 것일 수도 있으며, 유선 또는 무선 통신 네트워크를 통해 접속 가능한 전자 소프트웨어 유통망을 통하여 저장 또는 갱신된 것일 수도 있다. 저장부(105)는, 예를 들어, 주기억장치 및 보조기억장치 중 적어도 하나를 포함할 수 있고, 주기억장치는 롬(ROM) 및/또는 램(RAM)과 같은 반도체 저장 매체를 이용하여 구현 가능하며, 보조기억장치는 플래시 메모리 장치, SD(Secure Digital) 카드, 솔리드 스테이트 드라이브(SSD, Solid State Drive), 하드 디스크 드라이브(HDD) 또는 광 기록 매체(optical media) 등과 같이 데이터를 영구적 또는 반영구적으로 저장 가능한 적어도 하나의 저장 매체를 이용하여 구현 가능하다.

[0032] 출력부(109)는 카디널리티 추정 장치(100)의 동작과 관련된 데이터나, 카디널리티 추정 장치(100)의 상태나, 이용되는 프로그램 등을 외부로 출력할 수 있다. 예를 들어, 출력부(109)는 프로세서(110)의 처리 결과(일례로,

추정 대상 질의(20)과 추정 대상 질의(20)의 질의 대상이 되는 테이블(91 내지 99)에 해당하는 카디널리티의 추정 값(99)이나 학습용 데이터(10)에 의해 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등) 등을 시각적 또는 청각적으로 외부로 출력하여 사용자나 설계자 등에게 제공할 수도 있고, 이들 데이터나 학습 모델(113-1a, 113-2a, 113-3a, 115a 등) 등을 유무선 통신 네트워크 등을 통해 다른 정보처리장치(예를 들어, 추정 대상 질의(20)를 입력 받아 카디널리티 값을 추정할 다른 장치)로 전달할 수도 있다. 출력부(109)는, 실시예에 따라, 디스플레이, 프린터 장치, 스피커 장치, 영상 출력 단자, 데이터 입출력 단자 및/또는 통신 모듈 등을 포함할 수 있다. 출력부(109)는 카디널리티 주장 장치(100)와 일체형으로 마련된 것일 수도 있고 또는 임의적으로 분리 가능하게 마련된 것일 수도 있다.

[0033] 프로세서(110)는, 학습용 데이터(10)를 이용하여 적어도 하나의 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)을 훈련시키거나, 또는 이와 동시에 또는 이와 별도로 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등) 및 추정 대상 질의(20)를 이용하여 질의 대상이 되는 하나 또는 둘 이상의 테이블(91 내지 99)의 카디널리티 값을 적어도 하나 이상 추정할 수도 있다. 이 경우, 프로세서(110)는 질의(11, 20)에 대한 실제 카디널리티 값과, 훈련된 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)을 통해 추정된 카디널리티 값 사이의 오차가 최소화될 수 있도록 학습 모델(113-1a, 113-2a, 113-3a, 115a 등)을 훈련시킬 수 있다. 이를 위해 하기의 수학적 식 1과 같이 오차율(이하 추정 오차율)이 이용될 수 있다.

[0034] [수학적 식 1]

$$q - error = \max\left(\frac{act(q)}{est(q)}, \frac{est(q)}{act(q)}\right)$$

[0035]

[0036] 수학적 식 1에서 q-error는 추정 오차율이고, act(q)는 주어진 질의(q)에 대한 실제 카디널리티 값이며, est(q)는 주어진 질의(q)에 대한 추정된 카디널리티 값(도 4의 99)이다. 수학적 식 1은 그 연산 결과가 1과 동일하거나 1보다 크며, 실제 카디널리티 값과 추정 카디널리티 값이 동일하면 동일할수록(즉, 추정 결과가 실제 결과와 일치할수록) 수학적 식 1에 따른 연산 결과는 1에 근사하고, 실제 카디널리티 값과 추정 카디널리티 값의 차이가 상이하면 상이할수록(즉, 추정 결과가 실제 결과와 일치하지 않을수록) 1보다 상당히 크게 나타난다. 따라서, 프로세서(110)는 이와 같은 추정 오차율이 1에 가깝도록 학습 모델을 훈련시켜, 카디널리티 추정을 위한 훈련된 학습 모델을 획득할 수 있다.

[0037] 또한, 필요에 따라 프로세서(110)는 카디널리티 추정 장치(100)의 전반적인 동작을 제어하거나 또는 이의 동작에 필요한 각종 연산 처리 등을 수행할 수도 있다. 이러한 프로세서(110)의 동작(들)은 저장부(105)에 저장된 프로그램을 실행시켜 수행될 수도 있다. 프로세서(110)는, 예를 들어, 중앙 처리 장치(CPU: Central Processing Unit), 마이크로 컨트롤러 유닛(MCU: Micro Controller Unit), 애플리케이션 프로세서(AP: Application Processor), 전자 제어 유닛(ECU: Electronic Controlling Unit), 기본보드 관리 컨트롤러(BMC: Baseboard Management Controller), 마이컴(Micom: Micro Processor) 및/또는 이와 각종 연산 및 제어 처리를 수행할 수 있는 적어도 하나의 처리용 전자 장치 등을 포함할 수 있다. 이들 장치는, 예를 들어, 하나 또는 둘 이상의 반도체 칩, 회로 또는 관련 부품 등을 단독으로 이용하거나 조합하여 구현될 수 있다.

[0038] 도 1에 도시된 바를 참조하면, 프로세서(110)는 질의(11, 20)에 대한 벡터화를 수행하는 벡터화부(111)와, 질의(11, 20)의 대상이 되는 테이블(91 내지 99) 각각의 카디널리티 값의 추정을 위한 적어도 하나의 학습 모델(113-1a, 113-2a, 113-3a, 이하 제1 학습 모델)을 훈련하는 테이블처리부(113)와, 조인 질의 처리의 대상이 되는 모든 테이블(91 내지 99)에 관한 전체적인 카디널리티를 추정하거나 및/또는 이를 추정하기 위한 학습 모델(115a, 이하 제2 학습 모델)을 훈련시키는 종합처리부(115)를 포함할 수 있다.

[0039] 벡터화부(111)는, 학습용 데이터(10)의 학습용 질의(11)나 추정 대상 질의(20)를, 도 2 또는 도 3에 도시된 바와 같이 벡터의 형태로 변환하여 학습용 질의(11)나 추정 대상 질의(20)에 대응하는 소정의 벡터(이하 질의 벡터(91a 또는 90a))를 생성할 수 있다. 이 경우, 벡터화부(111)는, 하나의 테이블(91 내지 99 중 하나)에 대한 질의(11, 20)에 대응하는 하나 이상의 벡터(이하 제1 질의 벡터(91a))를 생성할 수도 있고, 및/또는 다수의 테이블(91 내지 99 중 둘 이상)에 대한 조인 질의(11, 20)에 대응하는 하나 또는 둘 이상의 벡터(90a, 이하 제2 질의 벡터)를 생성할 수도 있다. 제1 질의 벡터(91a)는 제1 질의 벡터(91a)에 상응하는 테이블처리부(113-1 내지 113-9)의 입력 값으로 이용될 수 있고, 제2 질의 벡터(90a)는 종합처리부(115)의 입력 값으로 이용될 수 있다. 여기서, 제1 질의 벡터(91a)에 상응하는 테이블처리부(113-1 내지 113-9)는 해당 제1 질의 벡터(91a)의 대상이 되는 적어도 하나의 테이블(91 내지 99) 각각에 대응하는 테이블처리부(113-1 내지 113-9)를 포함한다.

- [0040] 도 2는 각각의 테이블에 대응하는 학습 모델의 처리를 위한 조인 질의의 벡터화의 일례를 설명하기 위한 도면이다.
- [0041] 도 2에 도시된 바를 참조하면, 제1 질의 벡터(91a)는, 일 실시예에 있어서, 학습용 질의(11) 또는 추정 대상 질의(20)에서 이용된 연산자에 대한 정보(일례로 91-11 내지 91-13, 91-21 내지 91-23)를 포함할 수 있으며, 필요에 따라 학습용 질의(11) 또는 추정 대상 질의(20)에서 이용된 연산자에 통해 수행되는 연산에 이용된 값(숫자나 문자 등)을 포함하여 생성될 수 있다. 구체적으로 예를 들어, 제1 질의 벡터(91a)는 테이블(91 내지 99, 일례로 제1 테이블(91))의 모든 또는 일부의 속성(일례로 제1 속성(91-1) 및 제2 속성(91-2)) 각각에 대해서 적어도 4개의 요소(91-11 내지 91-14, 91-21 내지 91-24)를 포함하여 마련될 것일 수 있다. 이 경우, 어느 하나의 속성(일례로 제1 속성(91-1))의 요소(91-11 내지 91-14) 각각에는, 특정한 연산자(일례로 부등호(>, <)나 등호(=) 등의 비교 연산자)의 이용 여부에 대한 값(즉, 질의(11, 20)에서 이용된 연산자에 대한 정보)과, 특정 연산자에 따른 연산에 이용된 값(숫자나 문자 등)이 기록 및 설정될 수 있다. 보다 구체적으로는 제1 내지 제3 요소(91-11, 91-12, 91-13)에는 각각 질의(11, 20) 내에 오른쪽 부등호(>, Greater-than) 포함 여부, 질의(11, 20) 내 등호(=) 포함 여부 및 질의(11, 20) 내 왼쪽 부등호(<, Less than) 포함 여부에 대한 값이 순차적으로 기록되도록 마련될 수 있다. 이 경우, 질의(11, 20)가 오른쪽 부등호를 포함하는 경우에는 제1 요소(91-11)에는 1의 값이 기록 및 설정되고, 오른쪽 부등호를 포함하지 않는 경우에는 제1 요소(91-11)에는 0의 값이 설정될 수 있다. 마찬가지로 질의(11, 20)가 등호를 포함하는 경우에는 제2 요소(91-12)에는 1의 값이, 등호를 포함하지 않는 경우에는 0의 값이 기록될 수 있으며, 및/또는 제3 요소(91-13)에는 질의(11, 20)가 왼쪽 부등호를 포함하는 경우에는 1의 값이 기록되고 왼쪽 부등호를 포함하지 않는 경우에는 0의 값이 기록될 수 있다. 아울러 제4 요소(91-14)에는 연산자의 연산(일례로 비교 연산)에 사용된 값(일례로 수치) 등이 입력될 수 있다. 또한, 제2 속성(91-2)의 각각의 요소(91-21 내지 91-24)에 대해서도, 제1 속성(91-1)과 동일하게, 연산의 이용 여부에 대응하는 값(0 또는 1)이나 연산에 직접 사용된 수치 등이 기록될 수 있다. 실시예에 따라, 어느 하나의 속성(일례로 제2 속성(91-2))의 각각의 요소(91-21 내지 91-24)에 입력되는 값의 종류나 형태는, 다른 하나의 속성(일례로 제1 속성(91-1))의 각각의 요소(91-11 내지 91-14)에 입력되는 값의 종류나 형태와 상이할 수도 있다. 만약 조인 질의가 입력된 경우, 벡터화부(111)는 조인 질의에 대상이 되는 다수의 테이블(91 내지 99) 각각에 대해서 다수의 제1 질의 벡터(91a 등)를 생성할 수 있다. 즉, 적어도 하나의 테이블(91 내지 99 중 적어도 하나)를 대상으로 하는 질의(11, 20)에 대한 적어도 하나의 제1 질의 벡터와, 적어도 하나의 다른 테이블(91 내지 99 중 적어도 하나)를 대상으로 하는 질의(11, 20)에 대한 적어도 하나의 다른 제1 질의 벡터가 생성될 수 있다. 벡터화부(111)에 의해 생성된 하나 또는 둘 이상의 제1 질의 벡터(91a)는 테이블처리부(113)로 전달될 수 있다.
- [0042] 도 3은 다수의 테이블 간의 관계에 대한 학습 모델의 처리를 위한 조인 질의의 벡터화의 일례를 설명하기 위한 도면이다.
- [0043] 제2 질의 벡터(90a)는, 도 1 및 도 3에 에 도시된 바와 같이, 제1 질의 벡터(91a)의 각 속성(91-1, 91-2)에 해당하는 요소(들)(91-11 내지 91-14, 91-21 내지 91-24) 외에도 각 테이블(91 내지 99)의 실제 카디널리티 값이나 추정된 카디널리티 값(91b, 92b)을 더 포함할 수 있다. 예를 들어, 벡터화부(111)는 테이블처리부(113)로부터 제1 내지 제k 테이블처리부(113-1 내지 113-9) 각각의 의해 획득된 실제 또는 추정된 카디널리티 값을 수신하고, 이를 더 이용하여 제2 질의 벡터(90a)를 생성할 수 있다. 보다 구체적으로 예를 들어, 벡터화부(111)는, 제1 테이블처리부(113-1)의 출력 값인 카디널리티 값(91b, 실제 카디널리티 값일 수도 있고 추정된 카디널리티 값일 수도 있음. 이하 제1 카디널리티 값이라 함)을 제1 테이블처리부(113-1)로부터 획득하고, 이와 더불어 제1 테이블(91)의 각 속성(91-1, 91-2)에 해당하는 요소(들), 예를 들어, 학습용 질의(11) 또는 추정 대상 질의(20)에서 이용된 연산자에 대한 정보(91-11 내지 91-13, 91-21 내지 91-23)를 조합하여 제2 질의 벡터(90a)을 생성할 수 있다. 이 경우, 제2 질의 벡터(90a)는 연산에 이용된 값(숫자나 문자 등, 일례로 91-14, 92-24 등)은 포함하지 않고 생성될 수도 있다. 각 속성(91-1, 91-2)에 해당하는 요소(들)(91-11 내지 91-14, 91-21 내지 91-24)은 상술한 바와 동일하게 연산(일례로 부등호(>, <)나 등호(=) 등의 비교 연산)의 이용 여부에 상응하는 값이 기록될 수 있다. 여기서, 연산 이용 여부에 상응하는 값(91-11, 91-12, 91-13, 91-21, 91-22, 91-23)은, 예를 들어 상술한 바와 동일하게, 특정한 연산자의 사용 여부에 따라 0(사용하지 않음) 또는 1(사용함)의 값을 가질 수 있다.
- [0044] 또한, 제2 질의 벡터(90a)는, 일 실시예에 있어서, 모든 또는 일부의 테이블(일례로 제1 테이블(91) 및 제2 테이블(92)) 각각에 대응하는 영역(일례로 제1 영역(z1) 및 제2 영역(z2))을 하나 이상 포함할 수 있다. 이 경우, 벡터화부(111)는, 제1 카디널리티 값(91b) 및 제1 카디널리티 값(91b)에 대응하는 제1 테이블(91)의 각 속성(91-1, 91-2)에 해당하는 요소(들)(91-11 내지 91-13, 91-21 내지 91-23)을 조합하여 하나의 영역, 일례로 제1

영역(z_1)을 생성할 수 있다. 다시 말해서, 제1 영역(z_1)은, 질의(11, 20) 및 제1 테이블(91)에 대응하는 제1 카디널리티 값(91b)과, 어느 하나의 속성(일례로 제1 속성(91-1))에 대한 연산 사용 여부에 대한 값(91-11, 91-12, 91-13)과, 다른 하나의 속성(일례로 제2 속성(91-2))에 대한 연산 사용 여부에 대한 값(91-21, 91-22, 91-23) 등을 포함할 수 있다. 또한, 벡터화부(111)는 제2 테이블처리부(113-2)의 출력 값인 카디널리티 값(92b, 실제 카디널리티 값일 수도 있고 추정된 카디널리티 값일 수도 있음. 이하 제2 카디널리티 값이라 함) 등을 제2 테이블처리부(113-2)로부터 획득하고, 이와 더불어 각 속성(91-1, 91-2)에 해당하는 요소(들)(91-11 내지 91-14, 91-21 내지 91-24)을 조합하여 다른 하나의 영역, 일례로 제2 영역(z_2)을 생성할 수 있다. 따라서, 제2 영역(z_2)은 제1 영역(z_1)과 동일하게 질의(11, 20) 및 제2 테이블(92)에 대응하는 제2 카디널리티 값(92b)과, 제2 테이블(92)의 속성(일례로 제1 속성(92-1))에 대한 연산 사용 여부에 대한 값(92-11, 92-12, 92-13) 등을 포함할 수 있다. 제1 영역(z_1) 및 제2 영역(z_2)의 요소의 개수, 각 요소에 기록될 또는 기록된 값 및/또는 요소의 순서 등은 실시예나 상황에 따라서 동일할 수도 있고 상이할 수도 있다.

[0045] 일 실시예에 의하면, 제2 질의 벡터(90a)는, 질의(11, 20)에서 어떤 속성들이 조인을 하는데 이용되었는지 여부를 나타내는 부분(9, 이하 지시 부분)을 일 위치(예를 들어 마지막 요소)에 더 포함할 수도 있다. 지시 부분(9)은, 원-핫 인코딩(one-hot encoding) 방식을 기반으로 조인을 위해 이용된 속성을 나타내도록 마련된 것일 수도 있다. 지시 부분(9)은 하나 이상의 값(일례로 0 또는 1 등의 수치)를 단독으로 또는 조합하여 포함할 수 있다.

[0046] 벡터화부(111)에 의해 생성된 하나 또는 둘 이상의 제2 질의 벡터(90a)는 종합처리부(115)로 전달될 수 있다.

[0047] 도 4는 프로세서의 학습 동작의 일례를 설명하기 위한 도면이다.

[0048] 테이블처리부(113)는, 도 1 및 도 4에 도시된 바와 같이, 학습용 데이터(10)의 학습용 질의(11)에 대응하는 제1 질의 벡터(91a)를 입력 받고 이를 이용하여 각각의 테이블(91 내지 99) 중 적어도 하나에 대응하는 제1 학습 모델(113-1a, 113-2a, 113-3a)을 훈련시키거나, 및/또는 추정 대상 질의(20)에 대응하는 제1 질의 벡터(91a)를 기반으로 추정 대상 질의(20)의 대상이 되는 테이블(91 내지 99 중 적어도 하나)에 대한 출력 값(91b, 92b, 즉, 각 테이블(91 내지 92 중 적어도 하나)에 대한 카디널리티의 추정 값) 등을 획득할 수 있다. 일 실시예에 의하면, 테이블처리부(113)는 적어도 하나의 테이블(91 내지 99) 각각에 상응하는 하나 이상의 제1 내지 제k 테이블처리부(113-1 내지 113-9)를 포함할 수 있다. 각각의 제1 내지 제k 테이블처리부(113-1 내지 113-9)는, 도 4에 도시된 바와 같이, 하나 이상의 제1 학습 모델(113-1a, 113-2a, 113-3a)을 기반으로 구현될 수 있다. 여기서, 제1 학습 모델(113-1a, 113-2a, 113-3a)은, 제1 질의 벡터(91a)에 상응하여 적어도 하나의 테이블(91 내지 99) 각각에 대한 카디널리티의 추정 값을 획득할 수 있도록 훈련된 것일 수 있다. 제1 학습 모델(113-1a, 113-2a, 113-3a)은, 예를 들어, 다층 퍼셉트론(Multi-layer Perceptron), 심층 신경망(DNN: Deep Neural Network), convolutional 신경망(CNN: Convolutional Neural Network), 순환 신경망(RNN: Recurrent Neural Network), convolutional 순환 신경망(CRNN: Convolutional Recurrent Neural Network), 심층 신뢰 신경망(DBN: Deep Belief Network), 심층 Q-네트워크(Deep Q-Networks), 장단기 메모리(LSTM: Long short term memory), 서포트 벡터 머신(SVM: support vector machine), 생성적 적대 신경망(GAN: Generative Adversarial Network) 및/또는 조건적 생성적 적대 신경망(cGAN: Conditional GAN) 등의 적어도 하나의 학습 모델을 이용하여 구현될 수도 있으나, 이에 한정되는 것은 아니다. 예를 들어, 각각의 제1 내지 제k 테이블처리부(113-1 내지 113-9)의 제1 학습 모델(113-1a, 113-2a, 113-3a)은, 입력층, 2개의 은닉층 및 출력층을 갖도록 마련될 수도 있다. 각각의 제1 내지 제k 테이블처리부(113-1 내지 113-9)의 제1 학습 모델(113-1a, 113-2a, 113-3a)은 모두 동일할 수도 있고, 모두 상이할 수도 있으며, 일부는 동일하고 다른 일부는 상이할 수도 있다. 학습용 데이터(10)의 학습용 질의(11) 또는 추정 대상 질의(20)에 해당하는 제1 질의 벡터(91a)가 벡터화부(111)에 의해 생성되고 테이블처리부(113)에 입력되면, 제1 질의 벡터(91a)는, 해당 질의(11, 20)의 대상이 되는 테이블(일례로 제1 테이블(91))에 상응하는 테이블처리부(즉, 제1 테이블처리부(113-1))에 입력될 수 있다. 만약 제1 학습 모델(113-1a)에 대한 훈련이 수행되는 경우라면(즉, 제1 질의 벡터(91a)가 학습용 데이터(10)의 학습용 질의(11)로부터 획득된 경우라면), 제1 테이블처리부(113-1)의 제1 학습 모델(113-1a)은 입력된 제1 질의 벡터(91a)에 의해 훈련된다. 만약 추정 대상 질의(20)에 대한 카디널리티를 추정하는 경우라면(즉, 제1 질의 벡터(91a)가 추정 대상 질의(20)로부터 획득된 경우라면), 제1 테이블(91)에 해당하는 추정된 카디널리티(91b)가 획득될 수 있다. 다수의 질의 벡터(91a 등)가 생성된 경우에는, 다수의 제1 질의 벡터(91a 등)는, 해당 질의(11, 20)의 대상이 되는 테이블(91 내지 99 중 둘 이상)에 상응하는 테이블처리부(113-1 내지 113-9)에 각각 동시에 또는 순차적으로 입력되고, 다수의 제1 질의 벡터(91a 등)를 수신한 각각의 테이블처리부(113-1 내지 113-9)의 제1 학습 모델(113-1a, 113-2a, 113-3a)은 이를 기반으로 훈련되거나 및/또는 해당하는 테이블(91 내지 99 중 둘 이상) 각각에 대한 카디널리티의 추정 값

(91b, 92b)을 출력할 수 있다.

- [0049] 종합처리부(115)는, 일 실시예에 따르면, 도 1 및 도 4에 도시된 바와 같이, 조인 질의 형태의 학습용 데이터(10)의 학습용 질의(11)로부터 도출된 제2 질의 벡터(90a)를 수신하고 이를 이용하여 조인 질의에 관한 카디널리티 추정을 위한 적어도 하나의 제2 학습 모델(115a)을 훈련시키거나, 및/또는 조인 질의 형태의 추정 대상 질의(20)로부터 도출된 제2 질의 벡터(91a)를 수신하고 이를 기반으로 추정 대상 질의(20)의 에 대응하는 출력 값(99, 즉, 조인 질의에 대한 카디널리티의 추정 값)을 획득할 수 있다. 여기서, 제2 학습 모델(115a)은, 조인 질의에 관한 카디널리티의 추정을 위해 마련된 것으로, 예를 들어, 다층 퍼셉트론, 심층 신경망, 콘볼루션 신경망, 순환 신경망, 콘볼루션 순환 신경망, 심층 신뢰 신경망, 심층 Q-네트워크, 장단기 메모리, 서포트 벡터 머신, 생성적 적대 신경망 및 조건적 생성적 적대 신경망 중 적어도 하나를 포함할 수 있으나, 이에 한정되는 것은 아니다.
- [0050] 상술한 테이블처리부(113)에 의해 획득된 일반적인 질의(11, 20)에 대한 각 테이블(91 내지 99)의 카디널리티 추정 값(91b, 92b) 또는 상술한 테이블처리부(113) 및 종합처리부(115)를 통해 테이블(91 내지 99) 간의 상관 관계를 기반으로 획득된 조합 질의에 대한 카디널리티 추정 값(99)은, 하나 또는 둘 이상의 해당 테이블(91 내지 99)의 데이터 검색이나 자료 관리 등을 위해 이용될 수 있다. 여기서, 데이터 검색이나 자료 관리 등은 카디널리티 추정 장치(100)에 의해 수행될 수도 있고, 이와 연결된 다른 물리적 처리 장치에 의해 수행될 수도 있다.
- [0051] 이상 설명한 카디널리티 추정 장치(100)는, 일 실시예에 따르면, 상술한 학습용 데이터(질의 및/또는 카디널리티 등) 또는 추정 대상 질의의 입력과, 카디널리티 추정을 위한 학습 처리와, 이에 대응한 결과 획득 등을 수행할 수 있도록 특별히 고안된 하나 또는 둘 이상의 장치를 이용하여 구현될 수도 있다. 다른 실시예에 따르면, 카디널리티 추정 장치(100)는, 종래 알려진 적어도 하나의 정보처리장치를 단독으로 이용하거나 또는 조합 이용함으로써 구현될 수도 있다. 여기서, 적어도 하나의 정보처리장치는, 예를 들어, 서버용 하드웨어 장치, 데스크톱 컴퓨터, 랩톱 컴퓨터, 스마트 폰, 태블릿 피씨, 스마트 시계, 두부 장착형 디스플레이(HMD: Head Mounted Display) 장치, 내비게이션 장치, 휴대용 게임기, 디지털 텔레비전, 셋톱 박스, 디지털 미디어 재생 장치, 인공 지능 음향 재생 장치, 가전 기기(일례로 사물 인터넷을 이용하는 냉장고나 세탁기 등), 유인 이동체(일례로 승용차, 버스나 이륜차 등의 차량 등), 무인 이동체(일례로 로봇 청소기 등), 유인 비행체, 무인 비행체(드론 등), 산업용 기계, 가정용 또는 산업용 로봇 등을 포함할 수 있으나, 이에 한정되는 것은 아니다. 상황이나 필요에 따라서 상술한 정보처리장치 이외에도 정보의 연산 처리 및 제어가 가능한 다른 장치 중 적어도 하나가 상술한 카디널리티 추정 장치(10)로 채용될 수 있다.
- [0052] 이하 도 5를 참조하여 카디널리티 추정 방법의 일 실시예에 대해 설명하도록 한다.
- [0053] 도 5는 카디널리티 추정 방법에 대한 흐름도이다.
- [0054] 도 5에 도시된 바에 의하면 먼저 학습용 데이터가 획득될 수 있다(200). 학습용 데이터는 입력부 등을 통해 외부로부터 입력된 것일 수도 있고, 저장부 등에 기 저장된 것일 수도 있다. 학습용 데이터는 학습용 질의를 포함할 수 있으며, 학습용 질의는 하나의 테이블에 대한 질의를 포함할 수도 있고, 및/또는 다수의 테이블에 대한 질의(즉, 조인 질의)를 포함할 수도 있다. 또한, 학습용 데이터는 지도 학습을 위해 각각의 테이블에 대한 질의에 상응하는 실제 카디널리티 값이나, 다수의 테이블에 대한 조인 질의에 상응하는 실제 카디널리티 값을 더 포함할 수도 있다.
- [0055] 학습용 질의가 획득되면, 미리 정의된 바에 따라 또는 사용자 등의 수동 조작에 따라 학습용 질의에 대한 벡터화가 수행될 수 있다(202). 여기서, 학습용 질의에 대한 벡터화는, 각각의 테이블에 대한 학습용 질의에 따라 각각의 테이블에 대한 제1 질의 벡터를 획득하는 것을 포함할 수 있다. 제1 질의 벡터는, 질의에 포함된 소정의 연산자에 대한 정보를 포함할 수 있으며, 필요에 따라 소정 연산자에 의해 처리되는 값(수치 등) 등을 포함할 수 있다. 예를 들어, 제1 질의 벡터는 특정한 연산자(일례로 오른쪽 부등호, 등호 및 왼쪽 부등호 중 적어도 하나)의 존재 여부에 따라 0 또는 1의 값을 갖는 요소들을 순차적으로 포함할 수 있다.
- [0056] 제1 질의 벡터는 각각의 테이블에 대응하는 학습 모델에 입력되고, 각각의 학습 모델은 제1 질의 벡터에 의해 훈련된다(204). 이 경우, 각각의 테이블에 대한 질의에 상응하는 실제 카디널리티 값이 더 이용될 수도 있다. 여기서, 예를 들어, 학습 모델은 2개의 은닉층을 포함하여 구현될 수도 있다. 학습 모델은 다층 퍼셉트론, 심층 신경망, 콘볼루션 신경망, 순환 신경망, 콘볼루션 순환 신경망, 심층 신뢰 신경망, 심층 Q-네트워크, 장단기 메모리, 서포트 벡터 머신, 생성적 적대 신경망 또는 조건적 생성적 적대 신경망 등을 포함 가능하다. 각각의 테

이블에 대응하는 학습 모델은 상호 모두 동일할 수도 있고, 모두 상이할 수도 있으며, 일부는 동일하고 다른 일부는 상이할 수도 있다.

[0057] 조인 질의가 입력된 경우라면, 조인 질의에 대응하는 제2 질의 벡터가 제1 질의 벡터를 기반으로 한 학습 모델의 훈련에 따라 획득된 카디널리티 값(해당 테이블에 관한 실제 카디널리티 값 또는 해당 테이블에 관한 카디널리티의 추정 값을 포함할 수 있음)을 포함하여 생성되고, 생성된 제2 질의 벡터를 기반으로 조인 질의에 대응하는 카디널리티의 추정 값을 위한 학습 모델이 훈련될 수 있다(208). 제2 질의 벡터는, 다수의 테이블에 대응하는 영역을 하나 이상 포함할 수 있으며, 각각의 영역은 다수의 테이블 중 어느 하나에 대응하는 실제 또는 추정된 카디널리티 값과, 질의에 대한 정보를 포함할 수 있다. 예를 들어, 어느 하나의 영역은 해당 테이블의 실제 또는 추정된 카디널리티 값과, 해당 테이블을 질의 대상으로 삼는 질의에서 이용된 하나 이상의 특정한 연산자(일례로 오른쪽 부등호, 등호 및 왼쪽 부등호 중 적어도 하나)에 대한 정보(예를 들어, 특정한 연산자에 대한 정보는 특정한 연산자가 존재하는 경우에는 1로 설정되고, 부재하는 경우에는 0으로 설정될 수 있음)을 포함할 수 있다. 질의에서 이용된 특정 연산자 각각의 존재 여부에 대한 값은, 어느 하나의 영역 내에서 순차적으로 배치될 수도 있다. 또한, 제2 질의 벡터는 질의에서 어떤 속성들이 조인을 하는데 이용되었는지 여부를 나타내는 지시 부분을 더 포함할 수 있다. 지시 부분은 제2 질의 벡터의 맨 후단에 배치될 수 있으나, 이에 한정되는 것은 아니다. 지시 부분은 원-핫 인코딩 방식을 통해 질의에서 어떤 속성들이 조인에 이용되었는지를 나타내는 값(들)을 포함할 수 있다. 조인 질의의 학습을 위한 학습 모델은, 예를 들어, 3개의 은닉층을 포함할 수 있으나, 이에 한정되는 것은 아니다. 조인 질의의 학습을 위한 학습 모델은, 설계자나 사용자 등의 선택에 따라 다층 퍼셉트론, 심층 신경망, 콘볼루션 신경망, 순환 신경망, 콘볼루션 순환 신경망, 심층 신뢰 신경망, 심층 Q-네트워크, 장단기 메모리, 서포트 벡터 머신, 생성적 적대 신경망 또는 조건적 생성적 적대 신경망 등을 기반으로 구현될 수 있다. 질의의 학습을 위한 학습 모델은, 각각의 테이블에 대응하는 학습 모델의 전부 또는 일부와 동종의 모델을 기반으로 구현될 수도 있고, 또는 각각의 테이블에 대응하는 학습 모델과 이종의 모델을 기반으로 구현될 수도 있다.

[0058] 상술한 과정에 따라 학습 모델의 훈련이 수행되면, 순차적으로 또는 이와 동시에 추정 대상 질의가 획득될 수 있다(210). 추정 대상 질의는 학습용 데이터와 동일하거나 또는 상이한 방식으로 획득될 수 있다. 여기서, 추정 대상 질의에 대응하는 실제 카디널리티 값은 획득되지 않는다. 추정 대상 질의는, 하나의 테이블을 대상으로 하는 일반적인 질의일 수도 있고 또는 둘 이상의 테이블을 대상으로 하는 조인 질의일 수도 있다.

[0059] 추정 대상 질의에 대한 벡터화가 수행되고, 이에 대응하는 제1 질의 벡터가 획득될 수 있다(212). 추정 대상 질의에 대한 벡터화는 학습용 질의에 대한 벡터화와 동일하게 수행 가능하다. 예를 들어, 추정 대상 질의의 벡터화는 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 제1 질의 벡터를 획득하여 수행될 수도 있다. 제1 질의 벡터는, 상술한 바와 동일하게, 사용된 연산자에 대한 정보를 포함할 수 있으며, 사용된 연산자에 대한 정보는, 예를 들어, 특정한 연산자(일례로 오른쪽 부등호, 등호 및 왼쪽 부등호 중 적어도 하나)의 존재 여부에 따라 0 또는 1의 값을 갖는 요소들을 순차적으로 포함할 수 있다. 또한, 제1 질의 벡터는, 질의 내에서 사용된 연산자에 의해 처리되는 값 등을 더 포함할 수 있다.

[0060] 제1 질의 벡터는, 추정 대상 질의의 대상이 되는 테이블에 상응하는 학습 모델에 입력될 수 있다(214). 다수의 제1 질의 벡터가 획득된 경우, 다수의 제1 질의 벡터는, 각각 대응하는 다수의 학습 모델에 입력될 수 있다. 이에 따라 추정 대상 질의 및 추정 대상 질의의 대상이 되는 각각의 테이블에 대한 카디널리티의 추정 값이 획득될 수 있다.

[0061] 추정 대상 질의로 조인 질의가 입력된 경우라면, 하나 이상의 학습 모델이 각각 획득한 하나 이상의 카디널리티의 추정 값을 기반으로 조인 질의에 대응하는 제2 질의 벡터가 획득되고, 제2 질의 벡터를 또 다른 학습 모델에 입력하여 조인 질의에 대한 최종적인 카디널리티 추정 값이 획득될 수 있게 된다(216). 상술한 바와 동일하게 조인 질의인 추정 대상 질의에 관한 제2 질의 벡터는, 다수의 테이블에 대응하는 영역을 하나 이상 포함할 수 있으며, 각각의 영역은 다수의 테이블 중 어느 하나에 대응하는 카디널리티 추정 값과, 추정 대상 질의에 대한 정보를 포함할 수 있다. 추정 대상 질의에 대한 정보는, 하나 이상의 특정한 연산자(일례로 오른쪽 부등호, 등호 및 왼쪽 부등호 중 적어도 하나)의 이용 여부에 대한 정보(예를 들어, 존재 시에는 1로 설정되고, 부재 시에는 0으로 설정될 수 있음)을 포함할 수 있으며, 여기서 하나 이상의 특정한 연산자는 해당 영역에 상응하는 테이블을 질의 대상으로 삼는 질의에 포함된 것일 수 있다. 또한, 제2 질의 벡터는 지시 부분을 더 포함할 수도 있다. 지시 부분은 질의에서 어떤 속성들이 조인을 하는데 이용되었는지 여부를 일례로 원-핫 인코딩 등을 통해 나타낼 수 있다.

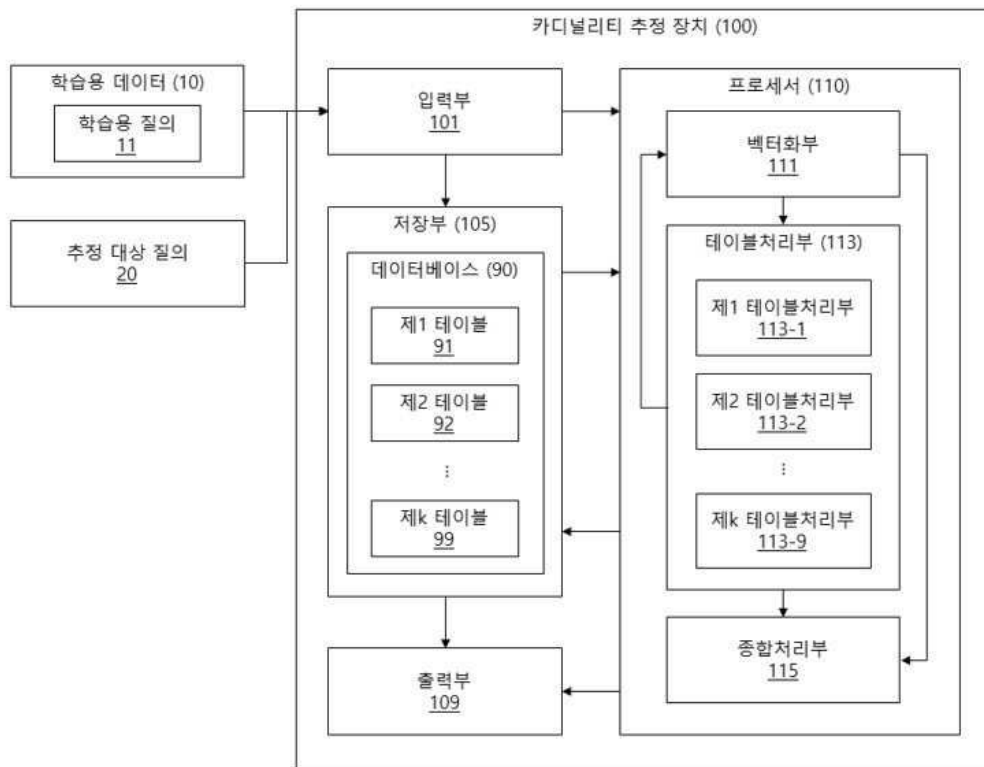
- [0062] 상술한 과정에 따라 획득된 조인 질의에 대한 카디널리티 추정 값은 조인 질의에 따른 다수의 테이블의 검색이나 유지 관리 등에 이용될 수 있다. 카디널리티 추정 값은, 학습 모델의 훈련 또는 카디널리티 추정 값을 획득한 소정의 장치에 의해 이용될 수도 있고, 소정 장치 이외의 다른 장치에 의해 이용될 수도 있다.
- [0063] 상술한 훈련 과정(200 내지 208) 및 카디널리티의 추정 과정(210 내지 216)은, 실시예에 따라서는, 동일한 장치에 의해 수행될 수도 있고 또는 상이한 장치에 의해 수행될 수도 있다. 상이한 장치에 의해 수행되는 경우, 어느 하나 이상의 장치의 처리에 의해 획득된 훈련된 학습 모델은 유무선 통신 네트워크나 휴대용 메모리 장치 등을 통해 다른 장치로 전달되고, 다른 장치는 훈련된 학습 모델을 기반으로 카디널리티의 추정을 수행하게 된다.
- [0064] 상술한 실시예에 따른 카디널리티 추정 방법은, 컴퓨터 장치에 의해 구동될 수 있는 프로그램의 형태로 구현될 수 있다. 프로그램은, 명령어, 라이브러리, 데이터 파일 및/또는 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 프로그램은 기계어 코드나 고급 언어 코드를 이용하여 설계 및 제작된 것일 수 있다. 프로그램은 상술한 방법을 구현하기 위하여 특별히 설계된 것일 수도 있고, 컴퓨터 소프트웨어 분야에서 통상의 기술자에게 공지되어 사용 가능한 각종 함수나 정의를 이용하여 구현된 것일 수도 있다. 또한, 여기서, 컴퓨터 장치는, 프로그램의 기능을 실현 가능하게 하는 프로세서나 메모리 등을 포함하여 구현된 것일 수 있으며, 필요에 따라 통신 장치를 더 포함할 수도 있다. 또한, 상술한 카디널리티 추정 방법을 구현하기 위한 프로그램은, 컴퓨터에 의해 판독 가능한 기록 매체에 기록될 수 있다. 컴퓨터에 의해 판독 가능한 기록 매체는, 예를 들어, 솔리드 스테이트 드라이브(SSD), 롬, 램 또는 플래시 메모리 등과 같은 반도체 저장 장치, 하드 디스크나 플로피 디스크 등과 같은 자기 디스크 저장 매체, 콤팩트 디스크나 디브이디 등과 같은 광 기록 매체, 플롭티컬 디스크 등과 같은 자기-광 기록 매체 및 자기 테이프 등 컴퓨터 등의 호출에 따라 실행되는 특정 프로그램을 저장 가능한 적어도 한 종류의 물리적 장치를 포함할 수 있다.
- [0065] 이상 카디널리티 추정 장치와 카디널리티 추정 방법의 여러 실시예에 대해 설명하였으나, 카디널리티 추정 장치 또는 카디널리티 추정 방법은 오직 상술한 실시예에 한정되는 것은 아니다. 해당 기술 분야에서 통상의 지식을 가진 자가 상술한 실시예를 기초로 수정 및 변형하여 구현 가능한 다양한 장치나 방법 역시 상술한 카디널리티 추정 장치 또는 카디널리티 추정 방법의 일 실시예가 될 수 있다. 예를 들어, 설명된 방법(들)이 설명된 바와 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치 또는 회로 등의 구성 요소(들)가 설명된 바와 다른 형태로 결합, 연결 또는 조합되거나 다른 구성 요소 또는 균등물 등에 의하여 대체/치환되더라도, 상술한 카디널리티 추정 장치 또는 카디널리티 추정 방법의 일 실시예가 될 수 있다.

부호의 설명

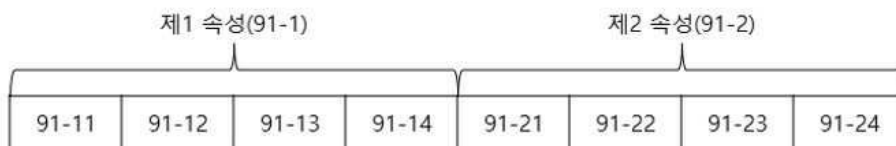
- [0066] 10: 학습용 데이터 11: 학습용 질의
20: 추정 대상 질의 100: 카디널리티 추정 장치
110: 프로세서 111: 벡터화부
113: 테이블처리부 115: 종합처리부

도면

도면1

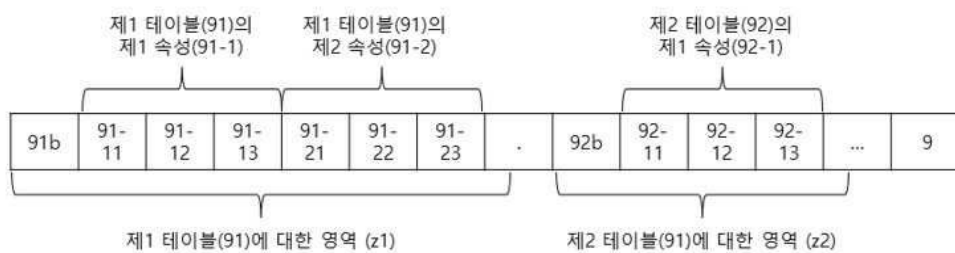


도면2



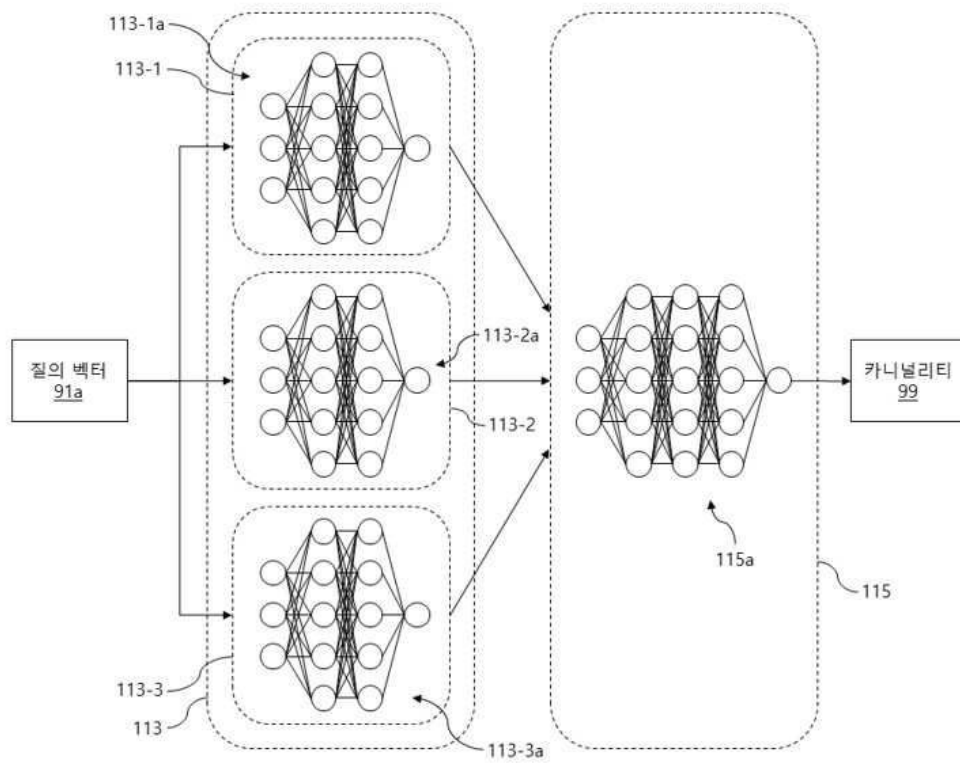
91a

도면3



90a

도면4



도면5

