(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0095005
(43) 공개일자 2025년06월26일

(51) 국제특허분류(Int. Cl.)
G06F 16/22 (2019.01) *G06F 16/901* (2019.01)
(52) CPC특허분류
G06F 16/2246 (2019.01)
G06F 16/9027 (2019.01)
(21) 출원번호 10-2023-0185603
(22) 출원일자 2023년12월19일
심사청구일자 2023년12월19일

(71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
신지선
서울특별시 강남구 선릉로126길 22, 103동 1202호 (삼성동, 롯데캐슬프레미어아파트)
카유모브 압두아지즈 이스맛 우글리
서울특별시 광진구 아차산로30길 20-8, 502호 (자양동, 로데빌)
(74) 대리인
특허법인린

전체 청구항 수 : 총 17 항

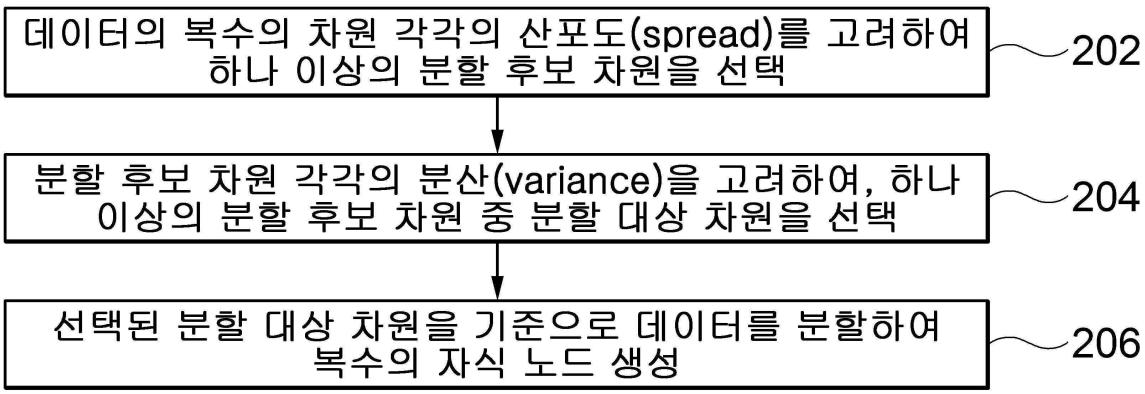
(54) 발명의 명칭 **SR-트리의 생성 방법 및 이를 수행하기 위한 컴퓨팅 장치**

(57) 요 약

SR-트리의 생성 방법 및 이를 수행하기 위한 컴퓨팅 장치가 개시된다. 일 실시예에 따른 방법은 각각 복수의 차원으로 구성되는 하나 이상의 데이터를 포함하는 임의의 노드에서, 복수의 차원 각각의 산포도(spread)를 고려하여 복수의 차원 중 하나 이상의 분할 후보 차원을 선택하는 단계; 하나 이상의 분할 후보 차원 각각의 분산(variance)을 고려하여, 하나 이상의 분할 후보 차원 중 분할 대상 차원을 선택하는 단계; 및 분할 대상 차원을 기준으로 하나 이상의 데이터를 분할하여 임의의 노드에 대한 복수의 자식 노드를 생성하는 단계를 포함한다.

대 표 도 - 도2

200



이 발명을 지원한 국가연구개발사업

과제고유번호	1711168648
과제번호	2020R1F1A1072275
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	무인항공기를 위한 블록체인 기반 보안 강화 기술 개발
기 여 율	1/2
과제수행기관명	세종대학교
연구기간	2022.03.01 ~ 2024.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1345366486
과제번호	2020R1A6A1A03038540
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	이공학학술연구기반구축(R&D)
연구과제명	자율지능무인비행체연구소
기 여 율	1/2
과제수행기관명	세종대학교
연구기간	2023.03.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로서,

각각 복수의 차원으로 구성되는 하나 이상의 데이터를 포함하는 임의의 노드에서, 상기 복수의 차원 각각의 산포도(spread)를 고려하여 상기 복수의 차원 중 하나 이상의 분할 후보 차원을 선택하는 단계;

상기 하나 이상의 분할 후보 차원 각각의 분산(variance)을 고려하여, 상기 하나 이상의 분할 후보 차원 중 분할 대상 차원을 선택하는 단계; 및

상기 분할 대상 차원을 기준으로 상기 하나 이상의 데이터를 분할하여 상기 임의의 노드에 대한 복수의 자식 노드를 생성하는 단계를 포함하는, SR-트리의 생성 방법.

청구항 2

청구항 1에 있어서,

상기 분할 후보 차원을 선택하는 단계는, 상기 복수의 차원 각각의 산포도가 큰 순서로 하나 이상의 차원을 상기 분할 후보 차원으로 선택하도록 구성되는, SR-트리의 생성 방법.

청구항 3

청구항 1에 있어서,

상기 분할 후보 차원을 선택하는 단계는, 산포도가 가장 큰 차원의 산포도를 기준으로 일정 범위 내의 산포도를 가지는 하나 이상의 차원을 상기 분할 후보 차원으로 선택하도록 구성되는, SR-트리의 생성 방법.

청구항 4

청구항 1에 있어서,

상기 분할 대상 차원을 선택하는 단계는, 상기 하나 이상의 분할 후보 차원 중 상기 분산이 가장 큰 차원을 상기 분할 대상 차원으로 선택하도록 구성되는, SR-트리의 생성 방법.

청구항 5

청구항 1에 있어서,

상기 분할 대상 차원을 선택하는 단계는, 상기 하나 이상의 분할 후보 차원 중, 산포도에 분산의 제곱근 값을 곱한 값이 가장 큰 차원을 상기 분할 대상 차원으로 선택하도록 구성되는, SR-트리의 생성 방법.

청구항 6

청구항 1에 있어서,

상기 산포도 및 상기 분산은, 상기 임의의 노드의 경계 구(sphere)의 중심으로부터 일정 거리 이상인 데이터만을 고려하여 계산되는, SR-트리의 생성 방법.

청구항 7

청구항 1에 있어서,

하나 이상의 분할 후보 차원을 선택하는 단계의 수행 전,

상기 임의의 노드 내의 데이터의 적어도 일부를 포함하는 i번째 경계 구($B_i = <c_i, r_i>$)에 대하여(여기서 c_i는 B_i의 중심, r_i는 B_i의 반지름), 상기 데이터 중 상기 중심(c_i)과 가장 먼 데이터(q_i)를 탐색하는 단계;

상기 q_i를 아웃라이어 세트에 추가하는 단계;

밑변이 $2*r_i$이고 높이(h_i)가 q_i와 c_i 사이의 거리인 이등변삼각형(abq_i)을 생성하는 단계;

상기 이등변삼각형의 각 꼭지점(a, b, q_i)과의 거리가 동일한 점을 새로운 구(B_{i+1})의 중심(c_{i+1})으로, 상기 중심(c_{i+1})에서 상기 각 꼭지점까지의 거리를 반지름(r_{i+1})으로 설정하는 단계; 및

상기 임의의 노드 내의 모든 데이터가 상기 경계 구에 포함될 때까지, 상기 탐색하는 단계 내지 설정하는 단계를 반복 수행하는 단계를 더 포함하는, SR-트리의 생성 방법.

청구항 8

청구항 7에 있어서,

상기 산포도 및 상기 분산은, 상기 아웃라이어 세트에 포함되는 데이터만을 고려하여 계산되는, SR-트리의 생성 방법.

청구항 9

하나 이상의 프로세서;

메모리; 및

하나 이상의 프로그램을 포함하는 장치로서,

상기 하나 이상의 프로그램은 상기 메모리에 저장되고 상기 하나 이상의 프로세서에 의해 실행되도록 구성되며,

상기 프로그램은,

각각 복수의 차원으로 구성되는 하나 이상의 데이터를 포함하는 임의의 노드에서, 상기 복수의 차원 각각의 산포도(spread)를 고려하여 상기 복수의 차원 중 하나 이상의 분할 후보 차원을 선택하는 단계;

상기 하나 이상의 분할 후보 차원 각각의 분산(variance)을 고려하여, 상기 하나 이상의 분할 후보 차원 중 분할 대상 차원을 선택하는 단계; 및

상기 분할 대상 차원을 기준으로 상기 하나 이상의 데이터를 분할하여 상기 임의의 노드에 대한 복수의 자식 노드를 생성하는 단계를 포함하는 단계들을 실행시키기 위한 명령어들을 포함하는, 컴퓨팅 장치.

청구항 10

청구항 9에 있어서,

상기 분할 후보 차원을 선택하는 단계는, 상기 복수의 차원 각각의 산포도가 큰 순서로 하나 이상의 차원을 상

기 분할 후보 차원으로 선택하도록 구성되는, 컴퓨팅 장치.

청구항 11

청구항 9에 있어서,

상기 분할 후보 차원을 선택하는 단계는, 산포도가 가장 큰 차원의 산포도를 기준으로 일정 범위 내의 산포도를 가지는 하나 이상의 차원을 상기 분할 후보 차원으로 선택하도록 구성되는, 컴퓨팅 장치.

청구항 12

청구항 9에 있어서,

상기 분할 대상 차원을 선택하는 단계는, 상기 하나 이상의 분할 후보 차원 중 상기 분산이 가장 큰 차원을 상기 분할 대상 차원으로 선택하도록 구성되는, 컴퓨팅 장치.

청구항 13

청구항 9에 있어서,

상기 분할 대상 차원을 선택하는 단계는, 상기 하나 이상의 분할 후보 차원 중, 산포도에 분산의 제곱근 값을 곱한 값이 가장 큰 차원을 상기 분할 대상 차원으로 선택하도록 구성되는, 컴퓨팅 장치.

청구항 14

청구항 9에 있어서,

상기 산포도 및 상기 분산은, 상기 임의의 노드의 경계 구(sphere)의 중심으로부터 일정 거리 이상인 데이터만을 고려하여 계산되는, 컴퓨팅 장치.

청구항 15

청구항 9에 있어서,

하나 이상의 분할 후보 차원을 선택하는 단계의 수행 전,

상기 임의의 노드 내의 데이터의 적어도 일부를 포함하는 i번째 경계 구($B_i = <c_i, r_i>$)에 대하여(여기서 c_i는 B_i의 중심, r_i는 B_i의 반지름), 상기 데이터 중 상기 중심(c_i)과 가장 먼 데이터(q_i)를 탐색하는 단계;

상기 q_i를 아웃라이어 세트에 추가하는 단계;

밑변이 $2*r_i$이고 높이(h_i)가 q_i와 c_i 사이의 거리인 이등변삼각형(abq_i)을 생성하는 단계;

상기 이등변삼각형의 각 꼭지점(a, b, q_i)과의 거리가 동일한 점을 새로운 구(B_{i+1})의 중심(c_{i+1})으로, 상기 중심(c_{i+1})에서 상기 각 꼭지점까지의 거리를 반지름(r_{i+1})으로 설정하는 단계; 및

상기 임의의 노드 내의 모든 데이터가 상기 경계 구에 포함될 때까지, 상기 탐색하는 단계 내지 설정하는 단계를 반복 수행하는 단계를 더 포함하는, 컴퓨팅 장치.

청구항 16

청구항 15에 있어서,

상기 산포도 및 상기 분산은, 상기 아웃라이어 세트에 포함되는 데이터만을 고려하여 계산되는, 컴퓨팅 장치.

청구항 17

비일시적 컴퓨터 판독 가능한 저장매체(non-transitory computer readable storage medium)에 저장된 컴퓨터 프로그램으로서,

상기 컴퓨터 프로그램은 하나 이상의 명령어들을 포함하고, 상기 명령어들은 하나 이상의 프로세서들을 갖는 컴퓨팅 장치에 의해 실행될 때, 상기 컴퓨팅 장치로 하여금,

각각 복수의 차원으로 구성되는 하나 이상의 데이터를 포함하는 임의의 노드에서, 상기 복수의 차원 각각의 산포도(spread)를 고려하여 상기 복수의 차원 중 하나 이상의 분할 후보 차원을 선택하는 단계;

상기 하나 이상의 분할 후보 차원 각각의 분산(variance)을 고려하여, 상기 하나 이상의 분할 후보 차원 중 분할 대상 차원을 선택하는 단계; 및

상기 분할 대상 차원을 기준으로 상기 하나 이상의 데이터를 분할하여 상기 임의의 노드에 대한 복수의 자식 노드를 생성하는 단계를 포함하는 단계들을 수행하도록 하는, 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 개시되는 실시예들은 SR-트리를 효율적으로 생성하기 위한 기술과 관련된다.

배 경 기 술

[0003] 다차원의 공간 데이터를 저장하는 색인(index)에는 여러 종류가 있다. 예를 들어, R-트리는 공간을 사각형(rectangle)로 분할하여 저장하고 SS-트리는 공간을 구(sphere)로 분할한다. SR-트리는 사각형과 구의 오버랩(intersection)으로 공간을 분할하며, 고차원 데이터에 주로 사용된다.

[0004] SR-트리에서 검색 성능을 높이기 위해서는 각 노드의 경계를 정하기 위한 구(sphere)의 크기를 최적으로 설정할 필요가 있다. 또한 주어진 정적 데이터에 대하여 SR-트리를 생성할 때 노드 간의 오버랩을 줄이기 위해서는 트리에 데이터를 순차적으로 추가하는 것이 아니라 데이터 전체를 이용하여 한번에 트리를 생성하는 이른바 벌크 로딩(bulk loading) 방식이 효과적이다. 그러나 기존의 벌크 로딩 방식은 데이터 중 일부 아웃라이어(outlier)의 영향으로 인해 전체 트리가 비효율적으로 생성되거나, 또는 트리 생성에 지나치게 많은 컴퓨팅 자원이 소모되는 문제가 있었다.

발명의 내용

해결하려는 과제

[0006] 개시되는 실시예들은 벌크 로딩 방식으로 효과적으로 SR-트리를 생성하기 위한 기술적인 수단을 제공하기 위한 것이다.

과제의 해결 수단

[0008] 예시적인 실시예에 따르면, 하나 이상의 프로세서들, 및 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로서, 각각 복수의 차원으로 구성되는

하나 이상의 데이터를 포함하는 임의의 노드에서, 복수의 차원 각각의 산포도(spread)를 고려하여 복수의 차원 중 하나 이상의 분할 후보 차원을 선택하는 단계; 하나 이상의 분할 후보 차원 각각의 분산(variance)을 고려하여, 하나 이상의 분할 후보 차원 중 분할 대상 차원을 선택하는 단계; 및 분할 대상 차원을 기준으로 하나 이상의 데이터를 분할하여 임의의 노드에 대한 복수의 자식 노드를 생성하는 단계를 포함하는, SR-트리의 생성 방법이 제공된다.

[0009] 일 실시예에서, 분할 후보 차원을 선택하는 단계는, 복수의 차원 각각의 산포도가 큰 순서로 하나 이상의 차원을 분할 후보 차원으로 선택하도록 구성될 수 있다.

[0011] 일 실시예에서, 분할 후보 차원을 선택하는 단계는, 산포도가 가장 큰 차원의 산포도를 기준으로 일정 범위 내의 산포도를 가지는 하나 이상의 차원을 분할 후보 차원으로 선택하도록 구성될 수 있다.

[0012] 일 실시예에서, 분할 대상 차원을 선택하는 단계는, 하나 이상의 분할 후보 차원 중 분산이 가장 큰 차원을 분할 대상 차원으로 선택하도록 구성될 수 있다.

[0013] 일 실시예에서, 분할 대상 차원을 선택하는 단계는, 하나 이상의 분할 후보 차원 중, 산포도에 분산의 제곱근 값을 곱한 값이 가장 큰 차원을 분할 대상 차원으로 선택하도록 구성될 수 있다.

[0014] 일 실시예에서, 산포도 및 분산은, 임의의 노드의 경계 구(sphere)의 중심으로부터 일정 거리 이상인 데이터만을 고려하여 계산될 수 있다.

[0015] 일 실시예에서, 상기 방법은 하나 이상의 분할 후보 차원을 선택하는 단계의 수행 전, 임의의 노드 내의 데이터의 적어도 일부를 포함하는 i번째 경계 구($B_i = <c_i, r_i>$)에 대하여(여기서 c_i는 B_i의 중심, r_i는 B_i의 반지름), 데이터 중 중심(c_i)과 가장 먼 데이터(q_i)를 탐색하는 단계;q_i를 아웃라이어 세트에 추가하는 단계; 밑변이 $2*r_i$이고 높이(h_i)가 q_i와 c_i 사이의 거리인 이등변삼각형(abq_i)을 생성하는 단계; 이등변삼각형의 각 꼭지점(a, b, q_i)과의 거리가 동일한 점을 새로운 구(B_{i+1})의 중심(c_{i+1})으로, 중심(c_{i+1})에서 각 꼭지점까지의 거리를 반지름(r_{i+1})으로 설정하는 단계; 및 임의의 노드 내의 모든 데이터가 경계 구에 포함될 때까지, 탐색하는 단계 내지 설정하는 단계를 반복 수행하는 단계를 더 포함할 수 있다.

[0016] 일 실시예에서, 산포도 및 분산은, 아웃라이어 세트에 포함되는 데이터만을 고려하여 계산될 수 있다.

[0017] 다른 예시적인 실시예에 따르면, 하나 이상의 프로세서; 메모리; 및 하나 이상의 프로그램을 포함하는 장치로서, 하나 이상의 프로그램은 메모리에 저장되고 하나 이상의 프로세서에 의해 실행되도록 구성되며, 프로그램은, 각각 복수의 차원으로 구성되는 하나 이상의 데이터를 포함하는 임의의 노드에서, 복수의 차원 각각의 산포도(spread)를 고려하여 복수의 차원 중 하나 이상의 분할 후보 차원을 선택하는 단계; 하나 이상의 분할 후보 차원 각각의 분산(variance)을 고려하여, 하나 이상의 분할 후보 차원 중 분할 대상 차원을 선택하는 단계; 및 분할 대상 차원을 기준으로 하나 이상의 데이터를 분할하여 임의의 노드에 대한 복수의 자식 노드를 생성하는 단계를 포함하는 단계들을 실행시키기 위한 명령어들을 포함하는, 컴퓨팅 장치가 제공된다.

[0018] 일 실시예에서, 분할 후보 차원을 선택하는 단계는, 복수의 차원 각각의 산포도가 큰 순서로 하나 이상의 차원을 분할 후보 차원으로 선택하도록 구성될 수 있다.

[0019] 일 실시예에서, 분할 후보 차원을 선택하는 단계는, 산포도가 가장 큰 차원의 산포도를 기준으로 일정 범위 내의 산포도를 가지는 하나 이상의 차원을 분할 후보 차원으로 선택하도록 구성될 수 있다.

[0020] 일 실시예에서, 분할 대상 차원을 선택하는 단계는, 하나 이상의 분할 후보 차원 중 분산이 가장 큰 차원을 분할 대상 차원으로 선택하도록 구성될 수 있다.

[0021] 일 실시예에서, 분할 대상 차원을 선택하는 단계는, 하나 이상의 분할 후보 차원 중, 산포도에 분산의 제곱근 값을 곱한 값이 가장 큰 차원을 분할 대상 차원으로 선택하도록 구성될 수 있다.

[0022] 일 실시예에서, 산포도 및 분산은, 임의의 노드의 경계 구(sphere)의 중심으로부터 일정 거리 이상인 데이터만을 고려하여 계산될 수 있다.

[0023] 일 실시예에서, 상기 방법은 하나 이상의 분할 후보 차원을 선택하는 단계의 수행 전, 임의의 노드 내의 데이터의 적어도 일부를 포함하는 i번째 경계 구($B_i = <c_i, r_i>$)에 대하여(여기서 c_i는 B_i의 중심, r_i는 B_i의 반지름),

데이터 중 중심(c_i)과 가장 먼 데이터(q_i)를 탐색하는 단계;q_i를 아웃라이어 세트에 추가하는 단계; 밑변이 $2*r_i$이고 높이(h_i)가 q_i와 c_i 사이의 거리인 이등변삼각형(abq_i)을 생성하는 단계; 이등변삼각형의 각 꼭지점(a, b, q_i)과의 거리가 동일한 점을 새로운 구(B_{i+1})의 중심(c_{i+1})으로, 중심(c_{i+1})에서 각 꼭지점까지의 거리를 반지름(r_{i+1})으로 설정하는 단계; 및 임의의 노드 내의 모든 데이터가 경계 구에 포함될 때까지, 탐색하는 단계 내지 설정하는 단계를 반복 수행하는 단계를 더 포함할 수 있다.

[0024] 일 실시예에서, 산포도 및 분산은, 아웃라이어 세트에 포함되는 데이터만을 고려하여 계산될 수 있다.

[0025] 다른 예시적인 실시예에 따르면, 비일시적 컴퓨터 판독 가능한 저장매체(non-transitory computer readable storage medium)에 저장된 컴퓨터 프로그램으로서, 컴퓨터 프로그램은 하나 이상의 명령어들을 포함하고, 명령어들은 하나 이상의 프로세서들을 갖는 컴퓨팅 장치에 의해 실행될 때, 컴퓨팅 장치로 하여금, 각각 복수의 차원으로 구성되는 하나 이상의 데이터를 포함하는 임의의 노드에서, 복수의 차원 각각의 산포도(spread)를 고려하여 복수의 차원 중 하나 이상의 분할 후보 차원을 선택하는 단계; 하나 이상의 분할 후보 차원 각각의 분산(variance)을 고려하여, 하나 이상의 분할 후보 차원 중 분할 대상 차원을 선택하는 단계; 및 분할 대상 차원을 기준으로 하나 이상의 데이터를 분할하여 임의의 노드에 대한 복수의 자식 노드를 생성하는 단계를 포함하는 단계들을 수행하도록 하는, 컴퓨터 프로그램이 제공된다.

발명의 효과

[0027] 개시되는 실시예들에 따를 경우 데이터의 산포도(spread) 및 분산(variance)을 고려하여 SR-트리의 노드를 분할함으로써 효과적으로 SR-트리를 생성할 수 있으며, 노드에 포함된 데이터의 일부만을 고려하여 산포도 및 분산을 계산함으로써 SR-트리 생성에 필요한 컴퓨팅 자원을 감소시킬 수 있다.

도면의 간단한 설명

[0029] 도 1은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도

도 2는 일 실시예에 따른 SR-트리의 생성 방법(200)을 설명하기 위한 순서도

도 3은 일 실시예에 따른 노도의 경계 구 계산 과정을 설명하기 위한 예시도

발명을 실시하기 위한 구체적인 내용

[0030] 이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 이하의 상세한 설명은 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.

[0031] 본 발명의 실시예들을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다. 상세한 설명에서 사용되는 용어는 단지 본 발명의 실시예들을 기술하기 위한 것이며, 결코 제한적이어서는 안 된다. 명확하게 달리 사용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 본 설명에서, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.

[0033] 도 1은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도이다. 도시된 실시예에서, 각 컴포넌트들은 이하에 기술된 것 이외에 상이한 기능 및 능력을 가질 수 있고, 이하에 기술되지 않은 것 이외에도 추가적인 컴포넌트를 포함할 수 있다.

[0034] 도시된 컴퓨팅 환경(10)은 컴퓨팅 장치(12)를 포함한다. 컴퓨팅 장치(12)는 적어도 하나의 프로세서(14), 컴퓨터 판독 가능 저장 매체(16) 및 통신 버스(18)를 포함한다. 프로세서(14)는 컴퓨팅 장치(12)로 하여금 앞서 언급된 예시적인 실시예에 따라 동작하도록 할 수 있다. 예컨대, 프로세서(14)는 컴퓨터 판독 가능 저장 매체(16)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 상기 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 상기 컴퓨터 실행 가능 명령어는 프로세서(14)에 의해 실행되는 경우 컴퓨팅 장치(12)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.

[0035] 컴퓨터 판독 가능 저장 매체(16)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능 저장 매체(16)에 저장된 프로그램(20)은 프로세서(14)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독 가능 저장 매체(16)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 컴퓨팅 장치(12)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.

[0036] 통신 버스(18)는 프로세서(14), 컴퓨터 판독 가능 저장 매체(16)를 포함하여 컴퓨팅 장치(12)의 다른 다양한 컴포넌트들을 상호 연결한다.

[0037] 컴퓨팅 장치(12)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(22) 및 하나 이상의 네트워크 통신 인터페이스(26)를 포함할 수 있다. 입출력 인터페이스(22) 및 네트워크 통신 인터페이스(26)는 통신 버스(18)에 연결된다. 입출력 장치(24)는 입출력 인터페이스(22)를 통해 컴퓨팅 장치(12)의 다른 컴포넌트들에 연결될 수 있다. 예시적인 입출력 장치(24)는 포인팅 장치(마우스 또는 트랙패드 등), 키보드, 터치 입력 장치(터치패드 또는 터치스크린 등), 음성 또는 소리 입력 장치, 다양한 종류의 센서 장치 및/또는 촬영 장치와 같은 입력 장치, 및/또는 디스플레이 장치, 프린터, 스피커 및/또는 네트워크 카드와 같은 출력 장치를 포함할 수 있다. 예시적인 입출력 장치(24)는 컴퓨팅 장치(12)를 구성하는 일 컴포넌트로서 컴퓨팅 장치(12)의 내부에 포함될 수도 있고, 컴퓨팅 장치(12)와는 구별되는 별개의 장치로 컴퓨팅 장치(12)와 연결될 수도 있다.

[0039] 도 2는 일 실시예에 따른 SR-트리의 생성 방법(200)을 설명하기 위한 순서도이다. 일 실시예에 따른 SR-트리의 생성 방법(200)은 주어진 데이터를 순차적으로 트리에 추가하지 않고, 벌크 로딩(bulk loading) 방식으로 데이터를 한번에 추가하여 트리를 생성하는 방식을 의미한다. 도시된 방법은 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치, 예컨대 전술한 컴퓨팅 장치(12)에서 수행될 수 있다. 도시된 흐름도에서는 상기 방법 또는 과정을 복수 개의 단계로 나누어 기재하였으나, 적어도 일부의 단계들은 순서를 바꾸어 수행되거나, 다른 단계와 결합되어 함께 수행되거나, 생략되거나, 세부 단계들로 나뉘어 수행되거나, 또는 도시되지 않은 하나 이상의 단계가 부가되어 수행될 수 있다.

[0040] 단계 202에서, 컴퓨팅 장치(12)는 SR-트리를 구성하는 임의의 노드에 포함된 데이터를 분할하기 위한 하나 이상의 분할 후보 차원을 선택한다. SR-트리 생성의 시작 단계에서는 루트 노드(root node)에 모든 데이터가 저장되므로, 이 경우 상기 임의의 노드는 루트 노드일 수 있다. 만약 SR-트리가 이미 복수의 노드를 가지고 있는 경우, 상기 임의의 노드는 해당 SR-트리에서 추가적으로 분할(split)하고자 하는 리프 노드(leaf node)일 수 있다.

[0041] 일 실시예에서, 데이터는 복수의 차원으로 구성되는 다차원 데이터일 수 있다. 본 단계에서, 컴퓨팅 장치(12)는 상기 복수의 차원 각각의 산포도(spread)를 고려하여 복수의 차원 중 하나 이상의 분할 후보 차원을 선택하도록 구성될 수 있다.

[0042] 예를 들어, 일 실시예에서 컴퓨팅 장치(12)는 복수의 차원 각각의 산포도가 큰 순서로 하나 이상의 차원(예를 들어, k개)을 분할 후보 차원으로 선택하도록 구성될 수 있다.

[0043] 다른 실시예에서, 컴퓨팅 장치(12)는 산포도가 가장 큰 차원의 산포도를 기준으로 일정 범위 내의 산포도를 가지는 하나 이상의 차원을 상기 분할 후보 차원으로 선택하도록 구성된다. 예를 들어, 산포도가 가장 높은 차원의 산포도 값이 D일 경우, 컴퓨팅 장치(12)는 산포도가 D*(1-e) (e는 0과 1 사이의 소수) 범위 내인 차원들을 상기 분할 후보 차원으로 선택할 수 있다.

[0045] 단계 204에서, 컴퓨팅 장치(12)는 단계 202에서 선택된 하나 이상의 분할 후보 차원 중 하나를 분할 대상 차원으로 선택한다.

[0046] 개시되는 실시예들에서, 컴퓨팅 장치(12)는 하나 이상의 분할 후보 차원 각각의 분산(variance)을 고려하여 분할 대상 차원을 선택할 수 있다.

[0047] 예를 들어, 일 실시예에서 컴퓨팅 장치(12)는 단계 202에서 선택된 하나 이상의 분할 후보 차원 중 분산이 가장 큰 차원을 분할 대상 차원으로 선택할 수 있다.

[0048] 다른 실시예에서, 컴퓨팅 장치(12)는 단계 202에서 선택된 하나 이상의 분할 후보 차원 중 산포도에 분산의 제곱근 값을 곱한 값(spread*sqrt(variance))이 가장 큰 차원을 분할 대상 차원으로 선택할 수 있다.

[0049] 산포도는 노드 내에 데이터가 얼마나 흩어져 있는지를 나타내는 값으로서, 일반적으로는 산포도를 고려하여 분할 대상 차원을 선택하게 된다. 그러나 산포도의 경우 아웃라이어(outlier)의 영향을 크게 받는 단점이 있다. 예를 들어 노드 내의 한두 개의 데이터가 다른 데이터와 동떨어져 있을 경우 산포도 값이 실제 데이터의 분포를 제대로 반영하지 못하는 경우가 발생할 수 있다. 따라서 개시되는 실시예들에서와 같이 산포도와 함께 분산을 고려하여 분할 대상 차원을 선택할 경우 아웃라이어의 영향을 효과적으로 감소시킬 수 있다.

[0051] 단계 206에서, 컴퓨팅 장치(12)는 단계 204에서 선택된 분할 대상 차원을 기준으로 상기 임의의 노드 내의 하나 이상의 데이터를 분할하여 복수의 자식 노드를 생성한다.

[0052] 일 실시예에서, 컴퓨팅 장치(12)는 기 설정된 최소 팬아웃(min fanout, 특정 노드에서 갈라지는 자식 노드의 최소값) 및 최대 팬아웃(max fanout, 특정 노드에서 갈라지는 자식 노드의 최대값)에 따라 상기 복수의 자식 노드를 생성할 수 있다.

[0053] 또한, 실시예에 따라 컴퓨팅 장치(12)는 상기 단계 202 및 204에서 한번 선택된 분할 대상 차원을 이용하여 전체 SR-트리를 생성하거나, 각 노드 별로 분할 대상 차원을 매번 계산하고 이를 이용하여 각 노드를 분할하도록 구성될 수 있다.

[0055] 한편, 몇몇 실시예들에서 상기 산포도 및 분산은 임의의 노드의 경계 구(sphere)의 중심으로부터 일정 거리 이상인 데이터만을 고려하여 계산될 수 있다. 전술한 바와 같이 SR-트리의 각 노드는 사각형과 구의 오버랩으로 그 범위가 결정된다. 각 노드의 산포도 및 분산은 주로 SR-트리의 경계를 결정하는 구 (또는 사각형)의 외곽 부분에 위치하는 데이터들의 영향을 많이 받게 된다. 따라서, 만약 분할 대상 노드의 경계를 결정하는 구의 중심으로부터 일정 거리 이상(예를 들어, 구의 중심에서 데이터까지의 거리가 반지름의 70% 이상 등)인 노드만을 이용하여 산포도 및 분산을 계산할 경우 계산량을 효과적으로 감소시킬 수 있다.

[0056] 또한 몇몇 실시예들에서, 임의의 노드의 경계 구는 해당 노드의 아웃라이어를 고려한 근사 알고리즘을 이용하여 계산될 수 있다. 전술한 바와 같이 SR-트리의 각 노드는 사각형과 구의 오버랩으로 그 범위가 결정되므로, 구의 사이즈가 타이트하게 (다시 말해, 해당 노드의 데이터를 모두 포함하면서 구의 크기가 최소화되도록) 정해져야 할 필요가 있다. 일반적으로는 해당 노드를 구성하는 각각의 데이터들의 평균점을 원점으로 하고 평균점에서 가장 먼 곳에 위치한 점(데이터)까지의 거리를 반지름으로 결정한다. 그러나 이 경우 데이터의 분포에 따라 구의 사이즈가 불필요하게 커지는 문제점이 있다. 이를 개선하기 위하여, 개시되는 실시예에서는 다음과 같은 알고리즘을 이용하여 노드의 경계 구를 설정하도록 구성될 수 있다.

[0057] 도 3은 일 실시예에 따른 노도의 경계 구 계산 과정을 설명하기 위한 예시도이다.

[0058] 예를 들어, 노드 내의 데이터의 적어도 일부를 포함하는 구 $B_i = <c_i, r_i>$에 대하여(여기서 c_i는 B_i의 중심, r_i는 B_i의 반지름), 노드 내의 데이터 중 B_i의 중심과 가장 먼 데이터(아웃라이어)를 q_i라 하자. 이 경우 q_i를 포함하는 새로운 구 $B_{i+1} = <c_{i+1}, r_{i+1}>$ 은 다음과 같이 계산될 수 있다.

[0059] (1) q_i를 아웃라이어 세트에 추가

[0060] (2) 밑변이 $2*r_i$이고 높이(h_i)가 q_i와 c_i 사이의 거리인 이등변삼각형 abq_i를 생성

[0061] (3) 이등변삼각형의 각 꼭지점(a, b, q_i)과의 거리가 동일한 점을 새로운 구의 중심 c_{i+1}로, 해당 점에서 각 꼭지점까지의 거리를 반지름 r_{i+1}로 설정

[0062] 상술한 (1) 내지 (3) 단계를 노드 내 모든 데이터가 구에 포함될 때까지 반복함으로써 노드의 경계 구를 계산할 수 있으며, 이와 같은 방법을 사용할 경우 단순히 각 데이터의 평균점을 이용하는 경우와 비교하여 경계구를 타이트하게 생성할 수 있다.

[0063] 또한, 상기와 같은 과정을 통해 임의의 노드의 경계 구가 계산될 경우, 도 2에서 설명한 산포도 및 분산은 상기 계산 과정에서 생성되는 아웃라이어 세트에 포함되는 데이터만을 고려하여 계산될 수 있다. 전술한 바와 같이 각 노드의 산포도 및 분산은 주로 SR-트리의 경계를 결정하는 구 (또는 사각형)의 외곽 부분에 위치하는 데이터들에 의하여 결정된다. 또한, 상기 아웃라이어 세트는 경계 구를 계산하는 과정에서 구의 중심으로부터 가장 먼 데이터들의 집합이므로, 이를 이용하여 산포도 및 분산을 계산할 경우 계산량을 효과적으로 감소시킬 수 있다.

[0065] 이상에서 대표적인 실시예를 통하여 본 발명에 대하여 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 전술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

부호의 설명

[0067] 10: 컴퓨팅 환경

12: 컴퓨팅 장치

14: 프로세서

16: 컴퓨터 판독 가능 저장 매체

18: 통신 버스

20: 프로그램

22: 입출력 인터페이스

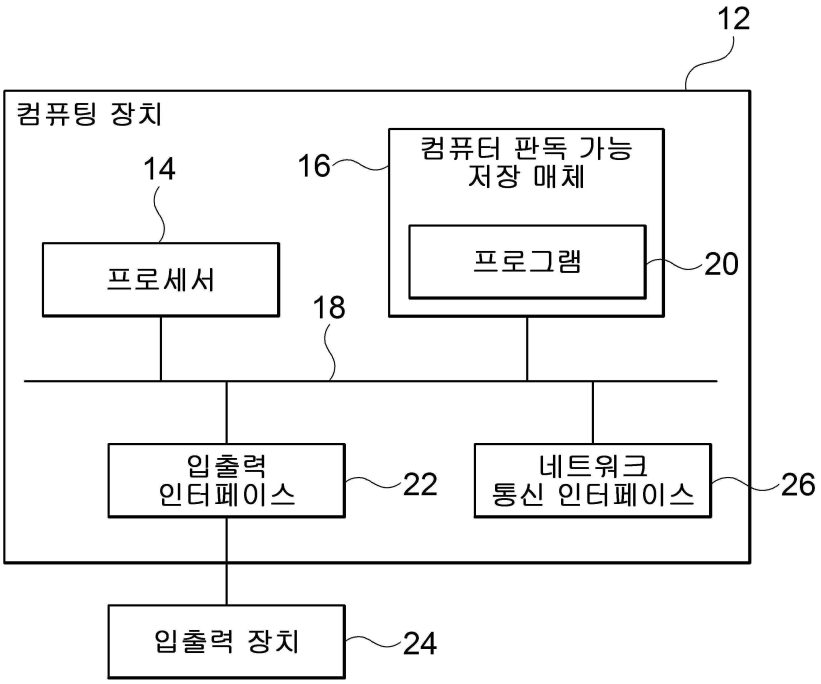
24: 입출력 장치

26: 네트워크 통신 인터페이스

도면

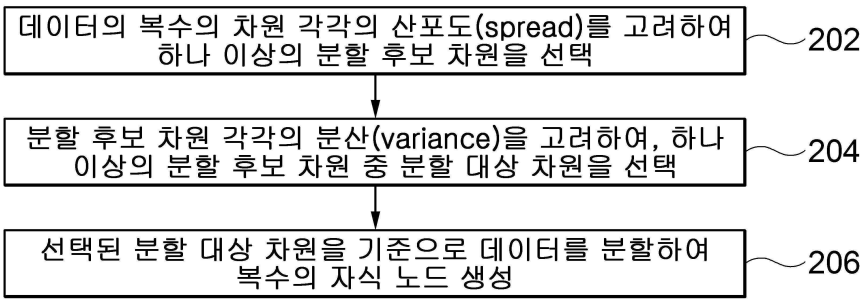
도면1

10



도면2

200



도면3

