

【서지사항】

【서류명】	특허출원서
【참조번호】	P20240563KR
【출원구분】	특허출원
【출원인】	
【명칭】	연세대학교 산학협력단
【특허고객번호】	2-2005-009509-9
【대리인】	
【명칭】	특허법인 지담
【대리인번호】	9-2018-100261-1
【지정된변리사】	김재흥
【포괄위임등록번호】	2023-020263-8
【발명의 국문명칭】	데이터 적응형 매개변수 조정을 통한 근사 최근접 이웃 검색 방법 및 프로그램
【발명의 영문명칭】	METHOD AND PROGRAM FOR APPROXIMATE NEAREST NEIGHBOR SEARCH VIA DATA-ADAPTIVE PARAMETER ADJUSTMENT
【발명자】	
【성명】	박상현
【성명의 영문표기】	PARK, Sang Hyun
【국적】	KR
【주민등록번호】	670101-1XXXXXX
【우편번호】	03626
【주소】	서울특별시 서대문구 세무서8길, 101동 1502호

【거주국】 KR

【발명자】

【성명】 진휘준

【성명의 영문표기】 JIN, Hui Jun

【국적】 CN

【주소】 서울특별시 신림동 490-28, 204호

【거주국】 KR

【발명자】

【성명】 이지은

【성명의 영문표기】 LEE, Ji Eun

【국적】 KR

【주민등록번호】 960912-2XXXXXX

【우편번호】 03716

【주소】 서울특별시 서대문구 동교로 291, 연희동임광아파트 101동
1004호

【거주국】 KR

【발명자】

【성명】 박성민

【성명의 영문표기】 PIA0, Sheng Min

【국적】 CN

【주소】 서울특별시 서대문구 연희로 25길 19, 301호

【거주국】 KR

【발명자】

【성명】 서상민

【성명의 영문표기】 SE0, Sang Min

【국적】 KR

【주민등록번호】 930507-1XXXXXX

【우편번호】 03724

【주소】 서울특별시 서대문구 연희로14길 29

【거주국】 KR

【발명자】

【성명】 권세인

【성명의 영문표기】 KWON, Se In

【국적】 KR

【주민등록번호】 970908-2XXXXXX

【우편번호】 03708

【주소】 서울특별시 서대문구 연희로 93, 연희오피스텔 702호

【거주국】 KR

【출원언어】 국어

【심사청구】 청구

【이 발명을 지원한 국가연구개발사업】

【과제고유번호】 2710008456

【과제번호】 I1170477

【부처명】 과학기술정보통신부

【과제관리(전문)기관명】 정보통신기획평가원

【연구사업명】 SW컴퓨팅산업원천기술개발

【연구과제명】 (SW 스타랩) IoT 환경을 위한 고성능 플래시 메모리 스토리
지 기반 인메모리 분산 DBMS 개발

【과제수행기관명】 연세대학교 산학협력단

【연구기간】 2024.01.01 ~ 2024.12.31

【취지】 위와 같이 특허청장에게 제출합니다.

대리인 특허법인 지담

(서명 또는 인)

【수수료】

【출원료】 0 면 46,000 원

【가산출원료】 48 면 0 원

【우선권주장료】 0 건 0 원

【심사청구료】 17 항 1,033,000 원

【합계】 1,079,000원

【감면사유】 전담조직(50%감면)[1]

【감면후 수수료】 539,500 원

【발명의 설명】

【발명의 명칭】

데이터 적응형 매개변수 조정을 통한 근사 최근접 이웃 검색 방법 및 프로그램
{METHOD AND PROGRAM FOR APPROXIMATE NEAREST NEIGHBOR SEARCH VIA DATA-
ADAPTIVE PARAMETER ADJUSTMENT}

【기술분야】

【0001】 본 발명은 데이터 검색 및 인덱싱 기술 분야에 관한 것으로, 특히 고차원 공간에서의 효율적인 근사 최근접 이웃(Approximate Nearest Neighbor; ANN) 검색을 수행하기 위한 방법 및 프로그램에 관한 것이다.

【발명의 배경이 되는 기술】

【0003】 최근접 이웃(Nearest Neighbor, NN) 검색은 추천 시스템, 컴퓨터 비전, 자연어 처리와 같은 다양한 분야에서 필수적인 기술로, 대규모 데이터셋에서 효율적으로 유사한 항목을 검색하는 데 중요하다. 일반적으로 사용되는 K-최근접 이웃(K-Nearest Neighbor, KNN) 검색은 주어진 데이터 포인트와 가까운 이웃을 빠르게 정확하게 찾기 위해 널리 활용된다. 그러나 데이터셋의 크기와 차원이 증가하면서, 기존의 KNN 방식은 실시간 처리를 위한 계산 부담이 커지며 대규모 환경에서 사용이 어려워지는 한계가 있다.

【0004】 이와 같은 확장성 문제를 해결하기 위해, 다양한 근사 최근접 이웃 (Approximate Nearest Neighbor, ANN) 검색 알고리즘이 개발되었으며, 높은 리콜 성능을 유지하면서도 계산 부담을 줄이는 것을 목표로 하고 있다. 그중에서도 계층적 네비게이블 스몰 월드(Hierarchical Navigable Small World, HNSW) 알고리즘은 다중 계층의 그래프 구조를 사용하여 높은 리콜과 빠른 검색 속도의 균형을 유지하며, ANN 검색에서 효과적인 솔루션으로 주목받고 있다. HNSW는 데이터 포인트들을 계층적인 구조로 구성하고 각 계층에서 근접 이웃을 탐색함으로써 대규모 이미지 검색, 텍스트 마이닝, 맞춤형 콘텐츠 추천 등 다양한 분야에 널리 활용되고 있다.

【0005】 그러나 HNSW 알고리즘은 그래프 생성 시 설정하는 매개변수 M (각 노드당 최대 연결 수)과 ef (검색 중 후보 목록의 크기; 검색 깊이)를 고정하여 사용하기 때문에, 데이터셋의 밀도나 차원이 상이한 경우 성능이 제한될 수 있다. 예를 들어, 데이터 밀도가 높은 영역에서는 주요 이웃을 포착하기 위해 높은 M 과 ef 값이 유리하지만, 밀도가 낮은 영역에서는 메모리와 계산 자원을 절감하기 위해 낮은 값이 필요하다. 이러한 고정된 매개변수 설정 방식은 데이터의 지역적 특성에 적응하지 못해, 고밀도 영역에서는 주요 이웃을 놓치고 저밀도 영역에서는 자원 낭비를 초래하는 문제가 발생할 수 있다.

【0006】 도 1은 2차원 공간에서 데이터 밀도 분포를 보여준다. 그림에서 주황색 노드는 고밀도 영역을 나타내며, 반경 내에 더 많은 이웃이 존재하여 데이터가 밀집되어 있다. 반면, 파란색 노드와 녹색 노드는 각각 중간 밀도, 저밀도 영역을 나타낸다. 이처럼 데이터 밀도가 낮은 저밀도 영역에서는 이웃이 적어 데이터가

희박한 분포를 보인다. 이러한 밀도 차이는 ANN 검색 성능에 직접적인 영향을 미치며, 매개변수 설정이 지역적 밀도에 맞추어 조정되지 않으면 효율성이 떨어질 수 있다.

【0007】 이 문제는 HNSW뿐만 아니라 NSG(Navigating Spreading-out Graph), DiskANN, EFANNA 등 다른 그래프 기반 ANN 검색 알고리즘에서도 발생하며, 전역적으로 고정된 매개변수 설정으로 인해 다양한 데이터 특성에 맞추기 어려운 한계가 있다. 일부 알고리즘들은 학습된 인덱스나 적응형 종료 전략과 같은 적응적 메커니즘을 도입하였으나, 이는 복잡한 튜닝을 요구하거나 특정 데이터 분포에 특화되는 경우가 많아 범용성이 떨어진다. 이와 같은 배경에서, 데이터의 밀도 변화에 동적으로 적응하면서도 최소한의 수동 개입으로 사용할 수 있는 보다 유연하고 일반화된 접근 방식이 필요하다.

【발명의 내용】

【해결하고자 하는 과제】

【0009】 본 발명이 이루고자 하는 기술적 과제는 상기한 문제점을 해결하기 위해 데이터 적응형 매개변수 조정을 통한 근사 최근접 이웃 검색을 위한 그래프 구축 방법 및 프로그램을 제공하는 것이다.

【0010】 구체적으로, 본 발명은 기존 HNSW알고리즘에서 고정된 매개변수 설정으로 인한 한계를 극복하고, 데이터의 밀도와 차원 특성에 따라 매개변수를 동적

으로 조정하는 데이터 적응형 프레임워크를 제공하는 것이다.

【0011】 이를 위해, 랜덤 프로젝션 기반의 밀도 추정 기법을 활용하여 각 데이터 포인트의 지역 밀도를 계산하고, 고밀도 영역에서는 더 많은 이웃과의 최대 연결 수와 검색 깊이를 증가시켜 리콜 성능을 높이며, 저밀도 영역에서는 메모리 사용과 계산 자원을 절감하기 위해 최대 연결 수와 검색 깊이를 감소시키는 방법 및 프로그램을 제공하는 것이다.

【0012】 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

【과제의 해결 수단】

【0014】 상기 기술적 과제를 달성하기 위하여, 데이터 밀도와 차원 특성에 따라 동적으로 매개변수를 조정하여 근사 최근접 이웃 검색을 위한 그래프를 구축하는 방법에 있어서, (a) 다차원 벡터로 표현된 다수의 데이터 포인트로 구성된 데이터셋 및 초기 매개변수를 수신하는 수신 단계; (b) 수신한 상기 데이터셋의 각 데이터 포인트의 밀도를 계산하는 단계; (c) 상기 밀도의 변동성을 고려하여 상기 매개변수의 범위를 결정하는 단계; (d) 결정된 상기 매개변수의 범위에 기초하여 특정 데이터 포인트의 밀도에 따라 동적으로 상기 매개변수를 조정하는 단계; 및

(e) 상기 특정 데이터 포인트를 조정된 상기 매개변수를 사용하여 그래프에 삽입하는 동적 그래프 구축 단계; 를 포함하고, 상기 매개변수는 '최대 연결 수(M_q)' 및 '검색 깊이(ef_q)'를 포함하는 것인 근사 최근접 이웃 검색을 위한 그래프 구축 방법을 제공한다.

【0015】 본 발명의 실시예에 있어서, 상기 (b) 단계는, RP(Random Projection) 방법을 사용하여 상기 데이터셋을 저차원 공간으로 투영한 후, 저차원으로 투영된 데이터셋에서 각 데이터 포인트에 대해 KNN(K-Nearest-Neighbors) 방법을 사용하여 밀도를 계산하는 것일 수 있다.

【0016】 또한, (f) 상기 데이터셋의 차원 수를 고려하여 상기 검색 깊이의 기준 값(ef_{ref})을 다음 수학적 식 1을 통해 설정하는 단계; 를 더 포함하는 것일 수 있다.

【0017】 또한, 상기 (c) 단계는, 상기 밀도 변동성을 데이터 밀도의 통계 값인 평균(ρ_μ)과 표준 편차(ρ_σ)로 나타내고, 상기 '최대 연결 수(M_q)'의 하한 값과 상한 값을 각각 다음 수학적 식 2 및 3을 통해 결정하고, 상기 '검색 깊이(ef_q)'의 하한 값과 상한 값을 각각 다음 수학적 식 4 및 5를 통해 결정하는 것일 수 있다.

【0018】 또한, 상기 (d) 단계는, 상기 '최대 연결 수(M_q)'를 다음 수학적 식 6을 통해 동적으로 조정하는 것일 수 있다.

【0019】 또한, 상기 (d) 단계는, 상기 '검색 깊이(ef_q)'를 다음 수학적 식 7을 통해 동적으로 조정하는 것일 수 있다.

【0020】 또한, 상기 데이터셋 내에 모든 데이터 포인트에 대해 상기 (d) 단계 및 상기 (e) 단계를 반복하여 수행하는 단계; 를 더 포함하는 것일 수 있다.

【0021】 또한, 상기 그래프는 HNSW(Hierarchical Navigable Small World) 그래프일 수 있다.

【0022】 상기 기술적 과제를 달성하기 위하여, 데이터 밀도와 차원 특성에 따라 동적으로 매개변수를 조정하여 근사 최근접 이웃 검색을 위한 그래프를 구축하는 컴퓨터 프로그램에 있어서, (a) 다차원 벡터로 표현된 다수의 데이터 포인트로 구성된 데이터셋 및 초기 매개변수를 수신하는 수신 단계; (b) 수신한 상기 데이터셋의 각 데이터 포인트의 밀도를 계산하는 단계; (c) 상기 밀도의 변동성을 고려하여 상기 매개변수의 범위를 결정하는 단계; (d) 결정된 상기 매개변수의 범위에 기초하여 특정 데이터 포인트의 밀도에 따라 동적으로 상기 매개변수를 조정하는 단계; 및 (e) 상기 특정 데이터 포인트를 조정된 상기 매개변수를 사용하여 그래프에 삽입하는 동적 그래프 구축 단계; 를 포함하는 근사 최근접 이웃 검색을 위한 그래프를 구축하는 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고, 상기 매개변수는 '최대 연결 수(M_q)' 및 '검색 깊이(ef_q)'를 포함하는 것인 컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램을 제공한다.

【0023】본 발명의 실시예에 있어서, 상기 (b) 단계는, RP(Random Projection) 방법을 사용하여 상기 데이터셋을 저차원 공간으로 투영한 후, 저차원으로 투영된 데이터셋에서 각 데이터 포인트에 대해 KNN(K-Nearest-Neighbors) 방법을 사용하여 밀도를 계산하는 것일 수 있다.

【0024】또한, 상기 근사 최근접 이웃 검색을 위한 그래프를 구축하는 방법은 (f) 상기 데이터셋의 차원 수를 고려하여 상기 검색 깊이의 기준 값(ef_{ref})을 다음 수학적 식 1을 통해 설정하는 단계; 를 더 포함하는, 것일 수 있다

【0025】또한, 상기 (c) 단계는, 상기 밀도 변동성을 데이터 밀도의 통계 값인 평균(ρ_μ)과 표준 편차(ρ_σ)로 나타내고, 상기 '최대 연결 수(M_q)'의 하한 값과 상한 값을 각각 다음 수학적 식 2 및 3을 통해 결정하고, 상기 '검색 깊이(ef_q)'의 하한 값과 상한 값을 각각 다음 수학적 식 4 및 5를 통해 결정하는 것일 수 있다.

【0026】또한, 상기 (d) 단계는, 상기 '최대 연결 수(M_q)'를 다음 수학적 식 6을 통해 동적으로 조정하는 것일 수 있다.

【0027】또한, 상기 (d) 단계는, 상기 '검색 깊이(ef_q)'를 다음 수학적 식 7을 통해 동적으로 조정하는 것일 수 있다.

【0028】또한, 상기 근사 최근접 이웃 검색을 위한 그래프를 구축하는 방법은 상기 데이터셋 내에 모든 데이터 포인트에 대해 상기 (d) 단계 및 상기 (e) 단계를 반복하여 수행하는 단계; 를 더 포함하는 것일 수 있다.

【0029】 또한, 상기 그래프는 HNSW(Hierarchical Navigable Small World) 그래프인 것일 수 있다.

【발명의 효과】

【0031】 본 발명의 실시예에 따르면, 데이터 적응형 매개변수 조정을 통한 근사 최근접 이웃 검색을 위한 그래프 구축 방법 및 프로그램을 제공할 수 있다.

【0032】 또한, 본 발명은 기존 HNSW 알고리즘의 고정된 매개변수 설정이 갖는 한계를 극복하고 효율적이고 정확한 근사 최근접 이웃 검색을 가능하게 한다.

【0033】 구체적으로, 본 발명은 랜덤 프로젝션 기반 밀도 추정 기법을 사용하여 각 데이터 포인트의 지역 밀도를 계산하고, 이를 바탕으로 고밀도 영역에서는 최대 연결 수와 검색 깊이를 증가시켜 리콜 성능을 높일 수 있다. 반면, 저밀도 영역에서는 메모리 사용과 계산 자원을 절감하기 위해 최대 연결 수와 검색 깊이를 낮춤으로써, 자원을 효율적으로 활용하는 최적화된 그래프 구조를 생성할 수 있다.

【0034】 결과적으로 본 발명은 다양한 밀도 분포를 가진 데이터셋에서도 높은 검색 성능과 효율성을 유지하여, 대규모 데이터 환경에서도 빠르고 정확한 근사 최근접 이웃 검색을 제공할 수 있다. 이러한 특징은 이미지 검색, 추천 시스템, 자연어 처리 등 실시간 대용량 데이터 처리가 필요한 다양한 분야에 본 발명의 응용 가능성을 높여줄 수 있다.

【0035】 본 발명의 효과는 상기한 효과로 한정되는 것은 아니며, 본 발명의 설명 또는 청구범위에 기재된 발명의 구성으로부터 추론 가능한 모든 효과를 포함하는 것으로 이해되어야 한다.

【도면의 간단한 설명】

【0037】 도 1은 2차원 공간에서 데이터 밀도 분포를 보여준다.

도 2는 본 발명의 실시예에 따른 데이터 밀도와 차원 특성에 따라 동적으로 매개변수를 조정하여 근사 최근접 이웃 검색을 위한 그래프를 구축하는 방법의 흐름도이다.

도 3은 본 발명의 실시예에 따른 데이터셋별 빌드 시간 및 메모리 사용량 개선률을 나타낸다.

도 4는 본 발명의 실시예에 따른 데이터셋별 리콜 성능 비교를 나타낸다.

도 5는 본 발명의 실시예와 비교예의 다양한 M 및 ef 값에 따른 성능 비교를 나타낸다.

도 6은 본 발명의 실시예의 확장 계수와 조정 계수의 성능 영향 분석을 나타낸다.

도 7은 본 발명의 실시예에 따른 근사 최근접 이웃 검색을 위한 그래프 구축 방법을 구현하는 컴퓨팅 장치를 도시한다.

【발명을 실시하기 위한 구체적인 내용】

【0038】 이하에서는 첨부한 도면을 참조하여 본 발명을 설명하기로 한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 따라서 여기에서 설명하는 실시예로 한정되는 것은 아니다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

【0039】 명세서 전체에서, 어떤 부분이 다른 부분과 "연결(접속, 접촉, 결합)"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 부재를 사이에 두고 "간접적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 구비할 수 있다는 것을 의미한다.

【0040】 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

【0041】 본 명세서에서, "모듈"은 하드웨어, 소프트웨어 또는 펌웨어로 구성된 유닛을 포함하며, 예컨대 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호

호환적으로 사용될 수 있다. 모듈은 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수 있다. 예컨대 모듈은 ASIC(application-specific integrated circuit)으로 구성될 수 있다.

【0043】 이하 첨부된 도면을 참조하여 본 발명의 실시예를 상세히 설명하기로 한다.

【0045】 도 2는 본 발명의 실시예에 따른 데이터 밀도와 차원 특성에 따라 동적으로 매개변수를 조정하여 근사 최근접 이웃 검색을 위한 그래프를 구축하는 방법의 흐름도이다.

【0046】 근사 최근접 이웃 검색을 위한 그래프 구축 방법은 데이터셋 및 초기 매개변수 수신 단계(S110), 데이터셋의 차원 축소 및 밀도 계산 단계(S120), 참조 검색 깊이 설정 단계(S130), 매개변수의 범위 결정 단계(S140), 매개변수 동적 조정 단계(S150), 동적 그래프 구축 단계(S160), 반복 단계(S170) 및 최종 그래프 출력 단계(S180)를 포함할 수 있다.

【0047】 데이터셋 및 초기 매개변수 수신 단계(S110)는, 근사 최근접 이웃 검색을 위한 그래프를 구축하기 위해, 시스템이 다차원 벡터로 표현된 다수의 데이터 포인트로 구성된 데이터셋 및 초기 매개변수를 사용자로부터 수신하는 단계이다.

【0048】 예를 들면, 이미지 특징 벡터를 사용한 검색을 수행할 경우, 각 데이터 포인트는 이미지의 특징을 나타내는 벡터로 구성될 수 있다.

【0049】 하나의 예로, 데이터셋 D 는 $\{x_1, x_2, \dots, x_n\}$ 의 형태로 입력될 수 있으며, 여기서, x_i 는 각 데이터 포인트를 나타내고, n 은 데이터셋 내 데이터 포인트의 개수를 의미한다.

【0050】 또한, 사용자는 초기 매개변수로 초기 최대 연결 수(M_{init})와 초기 검색 깊이(ef_{init})를 설정할 수 있다.

【0051】 초기 최대 연결 수(M_{init})는 각 노드가 가질 수 있는 최대 연결 수를 의미하며, 데이터 포인트가 그래프 내에서 연결할 수 있는 이웃 노드의 최대 개수를 결정한다. 본 발명의 일 실시예에 따르면, 사용자가 M_{init} 값을 설정하지 않은 경우, 시스템은 기본값으로 16을 사용할 수 있다.

【0052】 초기 검색 깊이(ef_{init})는 검색 중 유지할 후보 목록의 크기를 나타내며, 검색 깊이를 조절하는 역할을 한다. 이는 ANN 검색에서 고려할 후보군의 수를 제어하여 검색 정확도와 계산 효율성에 영향을 미친다. 본 발명의 일 실시예에 따르면, 사용자가 ef_{init} 값을 설정하지 않은 경우, 시스템은 기본값으로 100을 사용할 수 있다.

【0053】 데이터셋의 차원 축소 및 밀도 계산 단계(S120)는 시스템이 수신한 데이터셋으로부터 차원을 축소하고 각 데이터 포인트의 밀도를 계산하는 단계이다.

【0054】 고차원 데이터셋에서 데이터 포인트의 밀도를 계산하는 것은 일반적으로 많은 계산 자원을 소모하며, 고차원에서는 데이터 포인트 간 거리가 멀어져 이른바 '차원의 저주' 문제가 발생하여 밀도 계산이 어려워질 수 있다.

【0055】 이를 해결하기 위해, 상기 단계(S120)에서, RP-KNN(Random Projection - K-Nearest Neighbors) 방법을 사용하여 데이터셋의 차원을 축소한 후 밀도를 계산할 수 있다.

【0056】 상기 RP-KNN 방법에 따른 차원 축소 및 밀도 계산 과정을 예를 들어 설명하면 아래와 같다.

【0057】 먼저, 데이터셋 $D = \{x_1, x_2, \dots, x_n\}$ 의 각 데이터 포인트 x_i 를 랜덤 프로젝션을 통해 저차원 공간으로 투영할 수 있다. 이 랜덤 프로젝션은 Johnson-Lindenstrauss 보조정리에 기반해, 고차원 공간에서의 두 데이터 포인트 간의 상대적 거리를 높은 확률로 유지하면서 차원을 축소하는 방법이다.

【0058】 이후, 저차원으로 투영된 데이터셋에서 각 데이터 포인트에 대해 KNN 검색을 수행하여, 특정 데이터 포인트 x_i 주변의 K개의 이웃을 식별할 수 있다. 이를 통해, 각 데이터 포인트가 이웃 데이터 포인트들과 얼마나 밀집되어 있는지를 확인할 수 있으며, 이 정보를 바탕으로 밀도를 계산할 수 있다. 예를 들면, x_i 의 K-최근접 이웃까지의 평균 거리가 짧으면 해당 데이터 포인트는 고밀도 영역에 있다고 판단할 수 있고, 평균 거리가 길면 저밀도 영역에 있다고 판단할 수 있다.

【0059】 따라서, 각 데이터 포인트의 밀도 정보 $\rho = \{\rho_1, \rho_2, \dots, \rho_n\}$ 는 이웃 데이터 포인트들과 거리에 따라 계산될 수 있고, 이를 통해 밀도 추정의 계산 복잡도를 줄이면서도 고차원 데이터셋에 대해 높은 정확성을 유지할 수 있다.

【0060】 참조 검색 깊이 설정 단계(S130)는 데이터셋의 차원 수를 고려하여 검색 깊이의 기준값을 설정하는 단계이다.

【0061】 데이터셋의 차원이 높아질수록 데이터 포인트들 간의 거리가 구별되기 어려워지는 차원의 저주 문제가 발생하여, 효과적인 검색이 어려워진다. 이를 해결하기 위해, 데이터 포인트들 간 거리를 효율적으로 계산할 수 있도록 검색 깊이를 조정할 필요가 있다.

【0062】 따라서, 상기 단계(S130)에서는, 데이터셋의 차원을 기반으로 참조 검색 깊이(ef_{ref})를 계산하여 설정하고, 참조 검색 깊이(ef_{ref})는 다음 수학적 식 1을 통해 계산될 수 있다.

【0063】 [수학적 식 1]

$$\text{【0064】 } ef_{ref} = \left\lfloor ef_{init} + \left(\frac{\dim(D)}{\alpha} \right)^2 \right\rfloor$$

【0065】 여기서, ef_{init} 는 초기 검색 깊이를 의미하며, $\dim(D)$ 는 데이터셋의 차원 수를 의미한다. 매개변수 α 는 조정 파라미터를 의미한다. 조정 파라미터는 검색 깊이 조정의 크기를 제어하여 데이터 차원에 따른 검색 깊이의 증가율을 조절하는 역할을 한다. $\lfloor \cdot \rfloor$ 는 가장 가까운 정수로 내림하는 연산 기호를 의미한다. 이는

알고리즘에서 파라미터가 반드시 정수로 처리되어야 하기 때문에 사용된다.

【0066】 매개변수의 범위 결정 단계(S140)는 시스템이 데이터셋의 밀도 변동성을 고려하여 각 매개변수의 범위를 결정하는 단계이다.

【0067】 먼저, 밀도 변동성은 데이터 밀도의 통계 값인 평균(ρ_μ)과 표준 편차(ρ_σ)를 사용하여 나타낼 수 있다. 이 통계값을 활용하여 각 데이터의 밀도 특성에 맞는 매개변수를 설정함으로써, 데이터셋의 밀도 특성에 유연하게 적응할 수 있다.

【0068】 먼저, M_σ 의 범위, 즉 하한 값과 상한 값은 각각 다음 수학적식 2 및 3을 통해 설정될 수 있다.

【0069】 [수학적식 2]

$$\text{【0070】 } M_{low} = \max \left(2, \left\lfloor M_{init} - M_{init} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda \right\rfloor \right)$$

【0071】 [수학적식 3]

$$\text{【0072】 } M_{high} = M_{init} + M_{init} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda$$

【0073】 여기서, M_{init} 는 사용자가 설정한 초기 최대 연결 수를 의미하며, ρ_σ 는 데이터셋 밀도의 표준편차, ρ_μ 는 데이터셋 밀도의 평균을 나타내며, $\frac{\rho_\sigma}{\rho_\mu}$ 는 변동계수로 데이터셋 내 밀도의 변동성을 나타낸다, λ 는 확장 계수로, 밀도 변동에 따른 매개변수의 민감도를 조절하는 역할을 한다.

【0074】 이와 같이 설정된 변동 계수는 데이터셋의 다양한 영역에서, 밀도 변동 정도를 고려하도록 하여, 최대 연결 수가 데이터 밀도에 따라 비례적으로 조정될 수 있게 할 수 있다. 또한, 확장 계수 λ 는 밀도 변화에 대한 매개변수 조정의 민감도를 조절하여, 밀도가 높은 영역에서는 더 많은 이웃과 연결을 가능하게 하여 검색 정확도를 높이고, 밀도가 낮은 영역에서는 과소 연결을 방지하도록 조정할 수 있다.

【0076】 검색 깊이(ef_q)의 범위, 즉 하한 값과 상한 값은 각각 다음 수학적식 4 및 수학적식 5를 통해 설정될 수 있다.

【0077】 [수학적식 4]

$$\text{【0078】 } ef_{low} = \max(10, \left\lfloor ef_{ref} - ef_{ref} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda \right\rfloor)$$

【0079】 [수학적식 5]

$$\text{【0080】 } ef_{high} = \left\lceil ef_{ref} + ef_{ref} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda \right\rceil$$

【0081】 여기서, ef_{ref} 는 참조 검색 깊이를 나타낸다.

【0082】 ef_q 의 하한 값을 설정함으로써, 밀도가 희박한 노드에서도 충분한 후보 검색이 이루어지도록 보장하여 리콜 성능을 유지할 수 있다. 또한, ef_q 의 상한 값을 설정하여, 밀도가 높은 노드에서는 더 깊은 검색을 수행함으로써 더 정확한

최근접 이웃을 찾을 수 있도록 할 수 있다.

【0083】 마찬가지로, 변동 계수는 데이터셋의 다양한 영역 간 밀도 변동 정도를 반영하여 검색 깊이가 데이터 밀도에 따라 비례적으로 조정되도록 하며, 확장 계수는 지역 밀도 변화에 매개변수 조정 민감도를 조절하는 역할을 한다. 예를 들어, 밀도 변동이 큰 데이터셋에서는 λ 값을 높여 고밀도와 저밀도 영역 간의 매개변수 차이를 강조하고, 밀도 변동이 작은 데이터셋에서는 λ 값을 낮추어 전체적으로 일정한 매개변수 설정을 유지할 수 있다.

【0084】 이와 같은 방식으로 설정된 M_q 와 ef_q 의 범위는 데이터셋의 밀도 특성에 맞춰 동적으로 조정될 수 있다.

【0085】 매개변수 동적 조정 단계(S150)에서는 시스템이 결정된 매개변수의 범위에 기초하여 각 데이터 포인트의 밀도에 맞춰 동적으로 매개변수를 조정할 수 있다.

【0086】 예를 들어, 최대 연결 수 M_q 는 다음 수학식 6을 통해 동적으로 조정될 수 있다. 이 방식은 데이터 포인트 x_q 의 밀도 ρ_q 가 높을수록 M_q 값을 높게 설정하여 고밀도 영역에서 충분한 연결을 확보할 수 있도록 한다.

【0087】 [수학식 6]

【0088】
$$M_q = M_{low} + \left(\frac{\rho_q - \rho_{min}}{\rho_{max} - \rho_{min}} \right) \times (M_{high} - M_{low})$$

【0089】여기서, M_{low} 는 M_q 의 하한 값, M_{high} 는 M_q 의 상한 값, ρ_{min} 는 데이터 밀도의 최솟값, ρ_{max} 는 데이터 밀도의 최댓값을 의미한다.

【0091】검색 깊이 ef_q 는 다음 수학식 7을 통해 동적으로 조정될 수 있으며, 고밀도 영역에서는 깊이 있는 검색을 수행하도록 ef_q 값을 높이고, 저밀도 영역에서는 계산 자원을 절약하기 위해 ef_q 값을 낮출 수 있다.

【0092】[수학식 7]

$$\text{【0093】 } ef_q = ef_{low} + \left(\frac{\rho_q - \rho_{min}}{\rho_{max} - \rho_{min}} \right) \times (ef_{high} - ef_{low})$$

【0094】여기서, ef_{low} 는 ef_q 의 하한 값, ef_{high} 는 ef_q 의 상한 값, ρ_{min} 는 데이터 밀도의 최솟값, ρ_{max} 는 데이터 밀도의 최댓값을 의미한다.

【0095】본 발명의 실시예에 따르면, 데이터 포인트에 대해 동적으로 설정된 M_q 와 ef_q 는 지역 밀도 특성에 최적화된 매개변수를 적용하여, 고밀도 영역에서는 높은 리콜을 보장하고, 저밀도 영역에서는 자원을 절감하는 효과를 제공할 수 있다.

【0097】동적 그래프 구축 단계(S160)에서는, 각 데이터 포인트 x_q 를 앞서 계산된 동적 매개변수 M_q 와 ef_q 를 사용하여 그래프에 삽입할 수 있다. 이 단계에서

각 데이터 포인트는 동적으로 조정된 최대 연결 수와 검색 깊이에 따라 최적의 이웃 노드와 연결될 수 있다.

【0098】 이와 같이 동적 매개변수를 반영한 그래프를 구축함으로써, 고밀도 영역에서는 충분한 연결성과 깊이 있는 검색을 제공하고, 저밀도 영역에서는 메모리와 계산 자원을 효율적으로 절약할 수 있다.

【0099】 반복 단계(S170)에서는, 데이터셋 D 내 모든 데이터 포인트에 대해 매개변수 동적 조정 단계(S150)와 동적 그래프 구축 단계(S160)를 반복하여 수행할 수 있다.

【0100】 즉, 각 데이터 포인트 x_q 에 대해 밀도에 맞는 최적의 매개변수 설정을 계산하고, 이를 기반으로 그래프에 삽입하는 과정을 데이터셋의 모든 데이터 포인트에 대해 반복하여 수행함으로써, 그래프 전체가 데이터의 밀도 특성에 적응하도록 구성할 수 있다.

【0101】 이와 같은 반복 단계를 통해, 데이터셋 내 모든 데이터 포인트가 각자의 지역적 특성에 맞춰 연결되고 탐색될 수 있어, 고차원 대규모 데이터셋에서도 효율적인 근사 최근접 이웃 검색(ANN) 성능을 확보할 수 있다.

【0103】 최종 그래프 출력 단계(S180)에서는 모든 데이터 포인트가 최적의 매개변수 M_q 와 ef_q 에 따라 그래프에 연결되므로, 데이터 밀도 특성에 최적화된 그래프 구조가 형성될 수 있다.

【0104】 최종적으로 생성된 그래프는 데이터셋 내 각 포인트의 밀도에 따라 연결성이 동적으로 조정된 형태를 갖추게 된다. 고밀도 영역에서는 높은 최대 연결수와 높은 검색 깊이를 통해 포인트 간 빠르고 정확한 검색이 가능하며, 저밀도 영역에서는 불필요한 연결을 줄이고 얇은 검색 깊이로 구성하여 메모리와 계산 자원을 절약할 수 있도록 설계될 수 있다.

【0105】 이러한 구조적 특성 덕분에, 최종 그래프는 다양한 밀도 분포를 가진 데이터셋에서도 일관된 검색 성능과 효율성을 제공할 수 있다. 완성된 HNSW 그래프는 파일로 저장하거나 메모리 상에 로드되어 이후 검색 작업에 사용될 수 있다. 예를 들어, 그래프는 디스크에 저장하여 필요 시 불러오거나, 실시간 검색 작업을 위해 메모리 캐시에 상주시킬 수 있다.

【0106】 이 그래프는 근사 최근접 이웃(ANN) 검색 작업에서 입력 데이터 포인트와 유사한 이웃을 효율적으로 찾는 데 활용될 수 있다. 예를 들어, 대규모 이미지 검색 시스템에서, 사용자는 이미지를 업로드하여 유사한 이미지를 빠르게 찾는 작업을 수행할 수 있다. 최종 HNSW 그래프는 이러한 유사 이미지 검색 시, 업로드된 이미지의 특징 벡터를 기준으로 가장 가까운 이웃을 찾아 추천하는 데 활용될 수 있다. 또한, 최종 그래프는 텍스트 마이닝, 추천 시스템, 자연어 처리(NLP) 등 다양한 분야에서 실시간으로 유사 항목을 검색하고 추천하는 데 사용될 수 있다.

【0108】 [실시예: 검증 및 성능 평가]

【0109】 본 발명의 일 실시예에 따른 검증은 DHNSW(Dynamic Hierarchical Navigable Small World) 알고리즘의 성능을 기존 HNSW와 비교 평가하기 위해 수행되었다. 실험은 빌드 시간(Build Time), 메모리 사용량(Memory Usage), 리콜(Recall) 세 가지 주요 지표를 중심으로 진행되었으며, 이는 동적 매개변수 조정이 기존 정적 매개변수 설정에 비해 성능을 얼마나 개선하는지 확인하기 위한 것이다.

【0110】 실험 환경은 12세대 인텔 i7-12700F CPU와 94GB RAM을 장착한 서버로 구성되었으며, 이는 대규모 데이터셋을 처리하기에 충분한 성능을 제공하여 DHNSW 알고리즘의 확장성을 평가할 수 있도록 하였다.

【0111】 또한, 성능 평가를 위해 다양한 특성을 가지는 네 가지 데이터셋을 사용하였다. 각 데이터셋은 밀도 분포와 차원 수가 상이하여 DHNSW 알고리즘의 성능을 폭넓은 환경에서 평가하는 데 적합하다. 사용된 데이터셋은 다음과 같다.

【0112】 MNIST: 저차원 공간에서 비교적 균일한 밀도를 나타내는 손글씨 이미지 데이터셋으로, 낮은 차원에서의 성능을 평가할 수 있다.

【0113】 GloVe100K: 텍스트 임베딩 데이터셋으로, 중간 차원의 밀도 분포를 평가하는 데 적합하다.

【0114】 SIFT1M: 고차원에서 복잡한 밀도 분포를 나타내는 이미지 특징 벡터 데이터셋으로, 고차원에서의 성능을 확인할 수 있다.

【0115】 GIST1M: 고차원의 이미지 데이터셋으로, 복잡한 고차원 분포에서의 성능을 평가하는 데 적합하다.

【0116】 각 데이터셋의 상세한 정보는 아래 [표 1]에 요약되어 있다. 이러한 다양한 밀도 분포와 차원 특성을 반영한 데이터셋을 통해 DHNSW 알고리즘이 밀도 및 차원 변화에 어떻게 대응하는지를 종합적으로 평가할 수 있다.

【0118】 【표 1】

데이터셋	차원	용량	데이터 포인트수
MNIST	784	52 MB	60,000
loVe100K	300	990 MB	100,000
IFT1M	128	550 MB	1,000,000
IST1M	960	5.37 GB	1,000,000

【0119】 본 발명의 실시예에서 사용된 성능 평가 지표는 빌드 시간, 메모리 사용량, 리콜의 세 가지로 구성된다.

【0120】 빌드 시간(Build Time)은 각 데이터셋에 대해 HNSW 그래프를 구축하는 데 소요되는 총 시간을 의미한다. 특히 대규모 데이터셋의 경우, 빌드 시간 단축은 실시간 애플리케이션에서 중요한 요소이다. DHNSW 알고리즘의 빌드 시간 감소는 계산 효율성 향상을 의미하며, 대규모 환경에서도 신속한 그래프 생성을 가능하게 한다.

【0121】 메모리 사용량은 그래프 구축 과정 중 발생하는 최대 메모리 소비량을 의미한다. 알고리즘이 메모리 자원을 얼마나 효율적으로 관리하는지를 나타내는 지표이다. DHNSW 알고리즘은 메모리 사용량을 줄임으로써 제한된 메모리 환경에서도 안정적으로 동작할 수 있도록 하며, 이는 다양한 컴퓨팅 환경에서의 배포 가능

성을 높인다. 특히, 메모리 절약은 메모리 용량이 제한된 장치에서 매우 유용하다.

【0122】 리콜은 알고리즘이 실제 최근접 이웃을 얼마나 정확히 식별했는지를 나타내는 지표로, 검색 성능의 정확성을 평가한다. 이 지표는 알고리즘이 상위-k 결과에서 얼마나 많은 관련 이웃을 포착했는지를 반영하며, DHNSW 알고리즘이 고밀도 영역에서도 높은 검색 품질을 유지하도록 설계되었음을 보여준다. 높은 리콜은 검색의 정확도가 높음을 의미하며, 특히 추천 시스템이나 대규모 데이터 분석에서 중요한 성능 지표이다.

【0123】 본 검증의 목적은 DHNSW 알고리즘의 동적 매개변수 조정 방식이 기존 HNSW의 정적 매개변수 설정의 한계를 어떻게 개선하는지를 평가하는 데 있다. DHNSW가 다양한 밀도와 차원을 가진 데이터셋에서도 유연하게 확장 가능하며, 이를 통해 빌드 시간, 메모리 사용량, 리콜 성능에서 개선된 결과를 보임으로써 실시간 애플리케이션에 적합한 효율적인 ANN 검색 솔루션임을 입증하고자 한다.

【0124】 DHNSW 알고리즘은 데이터셋의 밀도와 차원 변화에 따라 매개변수를 동적으로 조정함으로써 검색 속도와 자원 효율성을 모두 최적화할 수 있는 장점을 제공한다. 이는 특히 실시간 추천 시스템, 텍스트 마이닝, 대규모 이미지 검색과 같은 다양한 애플리케이션에서 높은 유용성을 지닌다.

【0126】 도 3 은 본 발명의 실시예에 따른 데이터셋별 빌드 시간 및 메모리 사용량 개선률을 나타내며, 도 4는 본 발명의 실시예에 따른 데이터셋별 리콜 성능

비교를 나타낸다.

【0127】 본 발명의 실시예에 대한 성능 검증은 본 발명에서 제안한 DHNSW(Dynamic Hierarchical Navigable Small World) 알고리즘(=본 발명의 실시예)과 기존의 HNSW 알고리즘(=바닐라 HNSW, 비교예)을 네 가지 벤치마크 데이터셋에서 비교하는 방식으로 이루어졌다. 초기 설정값은 $M=16$, $ef=100$, $\alpha=100$, λ 로 설정하였으며, 그 결과는 도 3과 도 4에 요약되어 있다.

【0128】 본 발명의 실시예는 그래프 구축 과정에서는 빌드 시간을 크게 단축하는 효과를 보였다. 특히, 데이터셋의 크기와 차원이 클수록, 본 발명의 실시예는 동적 매개변수 조정을 통해 밀도와 차원 변화에 유연하게 대응하여 중복 계산을 최소화하고 그래프 생성 과정을 최적화함으로써 비교예보다 성능이 크게 향상되었다.

【0129】 도 3에 나타난 바와 같이, 본 발명의 실시예는 SIFT1M과 GIST1M 같은 고차원 대규모 데이터셋에서 빌드 시간의 개선 폭이 특히 크게 나타났다. 예를 들어, GIST1M 데이터셋에서는 51.35%의 빌드 시간 감소를 달성하여 DHNSW 알고리즘의 확장 가능성과 효율성을 입증하였다. 이러한 성능 향상은 밀도가 높은 영역에서 깊이 있는 검색을 수행하고, 저밀도 영역에서는 불필요한 계산을 줄이는 방식으로 이루어질 수 있다.

【0130】 또한, 본 발명의 실시예는 메모리 사용량을 효율적으로 관리하여, 비교예인 바닐라 HNSW 대비 메모리 소비량을 감소시키는 장점을 보였다. 이는 다양한 밀도 분포를 가진 데이터셋에서 특히 유리하게 작용하며, 불필요한 메모리 할당을 방지하여 메모리 자원을 절약할 수 있다. 도 3에 나타난 바와 같이, 본 발명의

실시예는 모든 데이터셋에서 메모리 사용량을 줄였으며, 그 감소 폭은 데이터셋의 크기와 복잡성에 따라 7.41%에서 32.44%까지 다양하였다. 예를 들어, MNIST 데이터셋에서는 메모리 사용량이 32.44% 감소하였고, 고차원인 GIST1M 데이터셋에서도 17.57%의 메모리 절감 효과가 관찰되었다. 이러한 메모리 절감 효과는 자원이 제한된 환경에서 DHNSW의 활용성을 높이며, 대규모 시스템에 적합한 알고리즘으로 작용할 수 있다.

【0131】 리콜(Recall) 성능에 있어서는, 본 발명의 실시예는 대부분의 데이터셋에서 비교예인 바닐라 HNSW와 유사한 수준을 유지하였으나, 약간의 감소가 나타났다. 이는 DHNSW의 리소스 효율성을 중시한 설계로 인해 발생할 수 있는 트레이드오프(trade-off)로, 빌드 시간과 메모리 사용량에서의 성능 개선을 고려했을 때 수용 가능한 수준이다. 도 4에 나타난 바와 같이, 본 발명의 실시예는 MNIST, GloVe100K, SIFT1M 데이터셋에서 리콜 값이 95% 이상을 유지하여, 대부분의 실제 애플리케이션에서 충분히 높은 검색 정확도를 제공할 수 있음을 확인하였다.

【0132】 반면, 고차원 데이터셋인 GIST1M에서는 리콜 값이 비교예(HNSW)의 74.03%에서 본 발명의 실시예(DHNSW)에서는 68.17%로 약간 감소하였으나, 이는 대규모 데이터 인덱싱, 실시간 추천 시스템 등에서 계산 효율성을 우선시하는 환경에서 수용 가능한 수준이다. 이러한 경우, DHNSW의 빌드 시간 및 메모리 절감 이점이 더 중요한 요소로 작용할 수 있다.

【0133】 종합적으로, 본 발명의 실시예는 다양한 밀도 분포와 차원 특성을 가진 데이터셋에서 유연하게 확장 가능한 솔루션임을 입증하였다. 특히 메모리 절

감과 빌드 시간 단축이 중요한 실시간 애플리케이션 환경에서 DHNSW는 효과적인 성능을 제공할 수 있다. DHNSW는 데이터셋의 특성에 따라 동적으로 최적화된 매개변수를 조정함으로써, 검색 효율성을 극대화할 수 있는 강력한 알고리즘으로 평가될 수 있다.

【0135】 도 5는 본 발명의 실시예에 따른 DHNSW(Dynamic Hierarchical Navigable Small World) 알고리즘과 비교예인 기존 HNSW(바닐라 HNSW)의 다양한 초기 매개변수 M 및 ef 값에 따른 성능 비교를 나타내며, 도 6은 DHNSW의 확장 계수(λ)와 조정 계수(α)의 성능 영향을 분석한 결과를 보여준다. 이 실험은 DHNSW 알고리즘의 다양한 초기 매개변수 설정에 대한 강건성과 주요 하이퍼 파라미터의 성능 영향을 평가하기 위해 수행되었다.

【0136】 각 데이터셋에서 1,000개의 데이터 포인트로 구성된 대표적인 부분 집합을 사용하여 실험 비용을 줄이면서도 데이터셋의 주요 특성을 보존하였다. 본 평가는 DHNSW와 바닐라 HNSW 간의 성능을 비교하기 위해 두 주요 매개변수인 M 과 ef 값을 체계적으로 변화시키며 수행되었다. M 은 8에서 64까지 8씩 증가시켰고, ef 는 50에서 400까지 50씩 증가시켜 다양한 매개변수 설정에서 DHNSW의 성능 변화를 분석하였다.

【0137】 도 5의 (a)와 (d)에 나타난 바와 같이, DHNSW는 모든 데이터셋에서 바닐라 HNSW보다 일관되게 더 짧은 구축 시간을 보였다. 특히 GIST1M과 같은 고차원 데이터셋에서 DHNSW의 구축 시간이 약 33% 감소하여, DHNSW가 다양한 데이터 밀

도와 차원 변화에 효과적으로 적응하는 능력을 입증하였다.

【0138】 또한, 도 5의 (b)와 (e)에서 확인할 수 있듯이, DHNSW는 모든 매개변수 구성에서 바닐라 HNSW보다 적은 메모리를 사용하였다. 다양한 밀도 분포를 가진 데이터셋에서는 DHNSW가 지역 데이터 밀도에 맞추어 메모리 자원을 동적으로 조정하여, 메모리 사용량을 최대 30-35%까지 줄였다. 이러한 메모리 절감 효과는 DHNSW가 성능 저하 없이 자원을 효율적으로 관리할 수 있음을 보여준다.

【0139】 도 5의 (c)와 (f)에 나타난 리콜(Recall) 결과는, DHNSW가 대부분의 경우 바닐라 HNSW와 유사한 수준의 리콜을 유지함을 보여준다. 매개변수가 달라짐에 따라 약간의 리콜 변동이 있었으나, 대부분의 경우 DHNSW와 바닐라 HNSW 간 리콜 차이는 1% 미만으로 나타났다. 이는 DHNSW의 동적 조정이 검색의 정확도를 유지하는 데 효과적임을 의미하며, 특히 낮은 M 및 ef 값에서도 유사한 리콜 성능을 달성하여, 동적 조정이 검색 정확도에 큰 영향을 미치지 않음을 입증한다.

【0141】 한편, DHNSW의 성능을 조절할 수 있는 두 가지 주요 하이퍼파라미터인 확장 계수(λ)와 조정 계수(α)의 영향을 분석하였다. 확장 계수 λ 는 데이터 밀도에 따른 매개변수 조정의 민감도를 제어하며, 조정 계수 α 는 차원 변화에 따른 매개변수 조정의 민감도를 제어한다. 이 두 하이퍼파라미터는 DHNSW가 다양한 데이터셋 특성에 적응할 수 있도록 지원한다. 실험은 기본 매개변수 설정인 $M=16$ 과 $ef=100$ 에서 수행되었으며, 결과는 도 6에 나타나 있다.

【0142】 도 6의 (a)와 (b)에 따르면, 확장 계수 λ 는 DHNSW의 구축 시간과 메모리 사용량에 큰 영향을 미쳤다. λ 값을 증가시키면 알고리즘이 매개변수를 더 유연하게 조정할 수 있어, 고차원 데이터셋에서 명확한 성능 개선이 나타났다. 예를 들어, GIST1M 데이터셋에서 λ 값을 증가시키면 구축 시간과 메모리 사용량이 현저히 감소하였으며, 이는 DHNSW가 고차원 데이터의 처리 요구에 효과적으로 대응함을 보여준다. 도 6의 (c)에서 나타나듯이, 리콜 성능은 약간 감소하였으나, 구축 시간과 메모리 사용량 절감에 비해 수용 가능한 수준으로 평가될 수 있다.

【0143】 조정 계수 α 는 도 6의 (d), (e), (f)에 나타난 바와 같이 구축 시간과 메모리 사용량에 상대적으로 안정적인 영향을 미쳤다. 대부분의 데이터셋에서 α 값의 변화에 따른 큰 성능 차이는 없었으나, $\alpha=50$ 설정 시 일부 데이터셋에서 바닐라 HNSW 대비 다소 개선된 리콜 성능을 보였다. 이는 동적 적응이 특정 데이터 분포에 대해 검색 성능을 미세 조정할 수 있으며, 특정 설정에서 더 높은 리콜을 달성할 수 있음을 시사할 수 있다.

【0144】 도 5와 도 6의 결과는 DHNSW가 다양한 매개변수 및 하이퍼파라미터 설정에서 바닐라 HNSW 대비 일관된 성능 개선을 제공하며, 밀도 변화에 따른 동적 조정의 강건성을 입증한다. 특히, 확장 계수 λ 는 고차원 데이터셋에서 메모리 사용량과 구축 시간을 크게 줄이는 데 중요한 역할을 하며, 조정 계수 α 는 리콜 성능에 미세한 조정을 가능하게 할 수 있다.

【0146】 따라서, 본 발명은 데이터 적응형 매개변수 조정을 통한 근사 최근

접 이웃 검색을 위한 그래프 구축 방법 및 프로그램을 제공할 수 있다.

【0147】 또한, 본 발명은 기존 HNSW 알고리즘의 고정된 매개변수 설정이 갖는 한계를 극복하고 효율적이고 정확한 근사 최근접 이웃 검색을 가능하게 한다.

【0148】 구체적으로, 본 발명은 랜덤 프로젝션 기반 밀도 추정 기법을 사용하여 각 데이터 포인트의 지역 밀도를 계산하고, 이를 바탕으로 고밀도 영역에서는 최대 연결 수와 검색 깊이를 증가시켜 리콜 성능을 높일 수 있다. 반면, 저밀도 영역에서는 메모리 사용과 계산 자원을 절감하기 위해 최대 연결 수와 검색 깊이를 낮춤으로써, 자원을 효율적으로 활용하는 최적화된 그래프 구조를 생성할 수 있다.

【0149】 결과적으로 본 발명은 다양한 밀도 분포를 가진 데이터셋에서도 높은 검색 성능과 효율성을 유지하여, 대규모 데이터 환경에서도 빠르고 정확한 근사 최근접 이웃 검색을 제공할 수 있다. 이러한 특징은 이미지 검색, 추천 시스템, 자연어 처리 등 실시간 대용량 데이터 처리가 필요한 다양한 분야에 본 발명의 응용 가능성을 높여줄 수 있다.

【0151】 도 7은 본 발명의 실시예에 따른 근사 최근접 이웃 검색을 위한 그래프 구축 방법을 구현하는 컴퓨팅 장치를 도시한다.

【0152】 도 2에 의해 설명된 본 발명의 실시예는 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치(200)로 구현될 수 있다.

【0153】컴퓨팅 장치(200)는 프로세서(210), 메모리(220), 스토리지(230), 통신 인터페이스(240), 시스템 인터커넥트(250) 및 디스플레이(260)를 포함할 수 있다.

【0154】프로세서(210)은, CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 및 APU(Application Processing Unit)을 포함한다.

【0155】메모리(220)는 프로세서(210)와 상호작용하여 프로그램이 효율적으로 실행될 수 있도록 데이터를 저장하고 필요한 정보에 빠르게 접근할 수 있도록 하는 기능을 수행한다. 메모리(220)는 레지스터, 캐시 메모리, 주 메모리, 읽기 전용 메모리, 가상 메모리, 비휘발성 메모리 중 적어도 하나를 포함한다.

【0156】스토리지(230)는 데이터를 영구적으로 저장하고 관리하는 역할을 한다. 스토리지는 컴퓨팅 시스템이 꺼지거나 재부팅된 후에도 데이터를 보존하며, 운영 체제, 애플리케이션, 사용자 파일 등을 저장하는 데 사용된다. 스토리지(230)은, 하드 디스크 드라이브(HDD), 솔리드 스테이트 드라이브(SSD), 광학 디스크, 네트워크 스토리지 및 클라우드 스토리지 중 적어도 하나를 포함한다.

【0157】통신 인터페이스(240)는 컴퓨팅 시스템 내부 및 외부의 다양한 장치들 간에 데이터를 주고받기 위한 경로를 제공한다. 통신 인터페이스(240)는 USB(Universal Serial Bus), PCIe(Peripheral Component Interconnect Express), SATA(Serial ATA), Ethernet, Wi-Fi, Thunderbolt 및 HDMI(High-Definition

Multimedia Interface) 중 적어도 하나의 통신 방식을 지원할 수 있다.

【0158】 시스템 인터커넥트(250)는 컴퓨팅 시스템 내부에서 다양한 구성 요소들 간의 데이터와 신호를 주고받는 역할을 한다. 시스템 인터커넥트(250)는, 버스(Bus), 포인트-투-포인트(Point-to-Point), 크로스바 스위치(Crossbar Switch), 네트워크-온-칩(Network-on-Chip, NoC) 중 적어도 하나의 방식을 지원할 수 있다.

【0159】 디스플레이(260)는 컴퓨팅 시스템의 출력 장치로서, 사용자에게 시각적인 정보를 제공하는 기능을 수행한다.

【0160】 전술한 구성에 의하여, 본 발명의 실시예에 따른 프로그램은, 프로세서(210)에 의해 실행되는 명령어들에 기초하여 실행되며, 메모리(220) 또는 스토리지(230)에 저장될 수 있다.

【0162】 전술한 본 발명의 실시예에 따른 방법은 다양한 컴퓨터 구성요소를 통하여 실행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능한 기록매체에 기록될 수 있다. 컴퓨터 판독 가능한 기록매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능한 기록매체에 기록되는 프로그램 명령은 본 발명의 실시예를 위하여 특별히 설계되고 구성된 것이거나, 컴퓨터 소프트웨어 분야의 통상의 기술자에게 공지되어 사용가능한 것일 수 있다. 컴퓨터 판독 가능한 기록매체는 하드디스크, 플로피디스크, 자기테이프 등의 자기기록 매체, CD-ROM, DVD 등의 광기록 매체, 플

롭티컬디스크 등의 자기-광 매체, ROM, RAM, 플래시 메모리 등과 같이, 프로그램 명령을 저장하고 수행하도록 구성된 하드웨어를 포함한다. 프로그램 명령은, 컴파일러에 의해 만들어지는 기계어 코드, 인터프리터를 사용하여 컴퓨터에서 실행될 수 있는 고급언어 코드를 포함한다. 하드웨어는 본 발명에 따른 방법을 처리하기 위하여 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있고, 그 역도 마찬가지이다.

【0163】 본 발명의 실시예에 따른 방법은 프로그램 명령 형태로 전자장치에서 실행될 수 있다. 전자장치는 스마트폰이나 스마트패드 등의 휴대용 통신 장치, 컴퓨터 장치, 휴대용 멀티미디어 장치, 휴대용 의료 기기, 카메라, 웨어러블 장치, 가전 장치를 포함한다.

【0164】 본 발명의 실시예에 따른 방법은 컴퓨터 프로그램 제품에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 기록매체의 형태로, 또는 어플리케이션 스토어를 통해 온라인으로 배포될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.

【0165】 본 발명의 실시예에 따른 구성요소, 예컨대 모듈 또는 프로그램 각각은 단수 또는 복수의 서브 구성요소로 구성될 수 있으며, 이러한 서브 구성요소들 중 일부 서브 구성요소가 생략되거나, 또는 다른 서브 구성요소가 더 포함될 수

있다. 일부 구성요소들(모듈 또는 프로그램)은 하나의 개체로 통합되어, 통합되기 이전의 각각의 해당 구성요소에 의해 수행되는 기능을 동일 또는 유사하게 수행할 수 있다. 본 발명의 실시예에 따른 모듈, 프로그램 또는 다른 구성요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

【0166】 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

【0167】 본 발명의 범위는 후술하는 청구범위에 의하여 나타내어지며, 청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

【부호의 설명】

【0170】 S110: 데이터셋 및 초기 매개변수 수신 단계

S120: 데이터셋의 차원 축소 및 밀도 계산 단계

S130: 참조 검색 깊이 설정 단계

S140: 매개변수의 범위 결정 단계

S150: 매개변수 동적 조정 단계

S160: 동적 그래프 구축 단계

S170: 반복 단계

S180: 최종 그래프 출력 단계

210: 프로세서

220: 메모리

230: 스토리지

240: 커뮤니케이션 인터페이스

250: 시스템 인터랙트

260: 디스플레이

【청구범위】

【청구항 1】

데이터 밀도와 차원 특성에 따라 동적으로 매개변수를 조정하여 근사 최근접 이웃 검색을 위한 그래프를 구축하는 방법에 있어서,

(a) 다차원 벡터로 표현된 다수의 데이터 포인트로 구성된 데이터셋 및 초기 매개변수를 수신하는 수신 단계;

(b) 수신한 상기 데이터셋의 각 데이터 포인트의 밀도를 계산하는 단계;

(c) 상기 밀도의 변동성을 고려하여 상기 매개변수의 범위를 결정하는 단계;

(d) 결정된 상기 매개변수의 범위에 기초하여 특정 데이터 포인트의 밀도에 따라 동적으로 상기 매개변수를 조정하는 단계; 및

(e) 상기 특정 데이터 포인트를 조정된 상기 매개변수를 사용하여 그래프에 삽입하는 동적 그래프 구축 단계; 를 포함하고,

상기 매개변수는 '최대 연결 수(M_a)' 및 '검색 깊이(ef_a)'를 포함하는 것인

근사 최근접 이웃 검색을 위한 그래프 구축 방법.

【청구항 2】

제1항에 있어서,

상기 (b) 단계는,

RP(Random Projection) 방법을 사용하여 상기 데이터셋을 저차원 공간으로

투영한 후, 저차원으로 투영된 데이터셋에서 각 데이터 포인트에 대해 KNN(K-Nearest-Neighbors) 방법을 사용하여 밀도를 계산하는 것인

근사 최근접 이웃 검색을 위한 그래프 구축 방법.

【청구항 3】

제1항에 있어서

(f) 상기 데이터셋의 차원 수를 고려하여 상기 검색 깊이의 기준 값(ef_{ref})을 다음 수학적 식 1을 통해 설정하는 단계; 를 더 포함하는

근사 최근접 이웃 검색을 위한 그래프 구축 방법.

[수학적 식 1]

$$ef_{ref} = \left\lfloor ef_{init} + \left(\frac{\dim(D)}{\alpha} \right)^2 \right\rfloor$$

여기서, ef_{init} 는 초기 검색 깊이를 의미하며, $\dim(D)$ 는 데이터셋의 차원 수를 의미하며, 매개변수 α 는 조정 파라미터를 의미하며, $\lfloor \cdot \rfloor$ 는 가장 가까운 정수로 내림하는 연산 기호를 의미함.

【청구항 4】

제3항에 있어서

상기 (c) 단계는,

상기 밀도 변동성을 데이터 밀도의 통계 값인 평균(ρ_μ)과 표준 편차(ρ_σ)로 나타내고,

상기 '최대 연결 수(M_a)'의 하한 값과 상한 값을 각각 다음 수학적 식 2 및 3을 통해 결정하고,

상기 '검색 깊이(ef_a)'의 하한 값과 상한 값을 각각 다음 수학적 식 4 및 5를 통해 결정하는 것인

근사 최근접 이웃 검색을 위한 그래프 구축 방법.

[수학적 식 2]

$$M_{low} = \max \left(2, \left\lfloor M_{init} - M_{init} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda \right\rfloor \right)$$

[수학적 식 3]

$$M_{high} = M_{init} + M_{init} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda$$

여기서, M_{init} 는 사용자가 설정한 초기 최대 연결 수를 의미하며, ρ_σ 는 데이터셋 밀도의 표준편차, ρ_μ 는 데이터셋 밀도의 평균을 나타내며, $\frac{\rho_\sigma}{\rho_\mu}$ 는 변동 계수로 데이터셋 내 밀도의 변동성을 나타내며, λ 는 확장 계수로, 밀도 변동에 따른 매개변수의 민감도를 조절하는 역할을 함.

[수학적 식 4]

$$ef_{low} = \max \left(10, \left\lfloor ef_{ref} - ef_{ref} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda \right\rfloor \right)$$

[수학식 5]

$$ef_{high} = \left\lfloor ef_{ref} + ef_{ref} \times \frac{\rho_{\sigma}}{\rho_{\mu}} \times \lambda \right\rfloor$$

여기서, ef_{ref} 는 참조 검색 깊이를 나타냄.

【청구항 5】

제1항에 있어서,

상기 (d) 단계는,

상기 '최대 연결 수(M_q)'를 다음 수학식 6을 통해 동적으로 조정하는 것인

근사 최근접 이웃 검색을 위한 그래프 구축 방법.

[수학식 6]

$$M_q = M_{low} + \left(\frac{\rho_q - \rho_{min}}{\rho_{max} - \rho_{min}} \right) \times (M_{high} - M_{low})$$

여기서, M_{low} 는 M_q 의 하한 값, M_{high} 는 M_q 의 상한 값, ρ_{min} 는 데이터 밀도의 최솟값, ρ_{max} 는 데이터 밀도의 최댓값을 의미함.

【청구항 6】

제1항에 있어서

상기 (d) 단계는,

상기 '검색 깊이(ef_q)'를 다음 수학적 식 7을 통해 동적으로 조정하는 것인

근사 최근접 이웃 검색을 위한 그래프 구축 방법.

[수학적 식 7]

$$ef_q = ef_{low} + \left(\frac{\rho_q - \rho_{min}}{\rho_{max} - \rho_{min}} \right) \times (ef_{high} - ef_{low})$$

여기서, ef_{low} 는 ef_q 의 하한 값, ef_{high} 는 ef_q 의 상한 값, ρ_{min} 는 데이터 밀도의 최솟값, ρ_{max} 는 데이터 밀도의 최댓값을 의미함.

【청구항 7】

제1항에 있어서,

상기 데이터셋 내에 모든 데이터 포인트에 대해 상기 (d) 단계 및 상기 (e) 단계를 반복하여 수행하는 단계; 를 더 포함하는

근사 최근접 이웃 검색을 위한 그래프 구축 방법.

【청구항 8】

제1항에 있어서

상기 그래프는

HNSW(Hierarchical Navigable Small World) 그래프인 것인

근사 최근접 이웃 검색을 위한 그래프 구축 방법.

【청구항 9】

데이터 밀도와 차원 특성에 따라 동적으로 매개변수를 조정하여 근사 최근접 이웃 검색을 위한 그래프를 구축하는 컴퓨터 프로그램에 있어서,

(a) 다차원 벡터로 표현된 다수의 데이터 포인트로 구성된 데이터셋 및 초기 매개변수를 수신하는 수신 단계;

(b) 수신한 상기 데이터셋의 각 데이터 포인트의 밀도를 계산하는 단계;

(c) 상기 밀도의 변동성을 고려하여 상기 매개변수의 범위를 결정하는 단계;

(d) 결정된 상기 매개변수의 범위에 기초하여 특정 데이터 포인트의 밀도에 따라 동적으로 상기 매개변수를 조정하는 단계; 및

(e) 상기 특정 데이터 포인트를 조정된 상기 매개변수를 사용하여 그래프에 삽입하는 동적 그래프 구축 단계; 를 포함하는 근사 최근접 이웃 검색을 위한 그래프 구축 방법을 프로세서가 수행하도록 하기 위한 명령어를 포함하고,

상기 매개변수는 '최대 연결 수(M_q)' 및 '검색 깊이(ef_q)'를 포함하는 것인

컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램.

【청구항 10】

제9항에 있어서,

상기 (b) 단계는,

RP(Random Projection) 방법을 사용하여 상기 데이터셋을 저차원 공간으로 투영한 후, 저차원으로 투영된 데이터셋에서 각 데이터 포인트에 대해 KNN(K-Nearest-Neighbors) 방법을 사용하여 밀도를 계산하는 것인

컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램.

【청구항 11】

제9항에 있어서

상기 근사 최근접 이웃 검색을 위한 그래프 구축 방법은

(f) 상기 데이터셋의 차원 수를 고려하여 상기 검색 깊이의 기준 값(ef_{ref})을 다음 수학적 식 1을 통해 설정하는 단계; 를 더 포함하는,

컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램.

[수학적 식 1]

$$ef_{ref} = \left\lfloor ef_{init} + \left(\frac{\dim(D)}{\alpha} \right)^2 \right\rfloor$$

여기서, ef_{init} 는 초기 검색 깊이를 의미하며, $\dim(D)$ 는 데이터셋의 차원 수를 의미하며, 매개변수 α 는 조정 파라미터를 의미하며, $\lfloor \cdot \rfloor$ 는 가장 가까운 정수로 내림하는 연산 기호를 의미함.

【청구항 12】

제11항에 있어서

상기 (c) 단계는,

상기 밀도 변동성을 데이터 밀도의 통계 값인 평균(ρ_μ)과 표준 편차(ρ_σ)로 나타내고,

상기 '최대 연결 수(M_q)'의 하한 값과 상한 값을 각각 다음 수학적식 2 및 3을 통해 결정하고,

상기 '검색 깊이(ef_q)'의 하한 값과 상한 값을 각각 다음 수학적식 4 및 5를 통해 결정하는 것인

컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램.

[수학적식 2]

$$M_{low} = \max \left(2, \left\lfloor M_{init} - M_{init} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda \right\rfloor \right)$$

[수학적식 3]

$$M_{high} = M_{init} + M_{init} \times \frac{\rho_\sigma}{\rho_\mu} \times \lambda$$

여기서, M_{init} 는 사용자가 설정한 초기 최대 연결 수를 의미하며, ρ_σ 는 데이터셋 밀도의 표준편차, ρ_μ 는 데이터셋 밀도의 평균을 나타내며, $\frac{\rho_\sigma}{\rho_\mu}$ 는 변동 계수로 데이터셋 내 밀도의 변동성을 나타내며, λ 는 확장 계수로, 밀도 변동에 따른 매개변

수의 민감도를 조절하는 역할을 함.

[수학식 4]

$$ef_{low} = \max(10, \left\lfloor ef_{ref} - ef_{ref} \times \frac{\rho_{\sigma}}{\rho_{\mu}} \times \lambda \right\rfloor)$$

[수학식 5]

$$ef_{high} = \left\lceil ef_{ref} + ef_{ref} \times \frac{\rho_{\sigma}}{\rho_{\mu}} \times \lambda \right\rceil$$

여기서, ef_{ref} 는 참조 검색 깊이를 나타냄.

【청구항 13】

제9항에 있어서,

상기 (d) 단계는,

상기 '최대 연결 수(M_q)'를 다음 수학식 6을 통해 동적으로 조정하는 것인
컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램.

[수학식 6]

$$M_q = M_{low} + \left(\frac{\rho_q - \rho_{min}}{\rho_{max} - \rho_{min}} \right) \times (M_{high} - M_{low})$$

여기서, M_{low} 는 M_q 의 하한 값, M_{high} 는 M_q 의 상한 값, ρ_{min} 는 데이터 밀도의 최솟값, ρ_{max} 는 데이터 밀도의 최댓값을 의미함.

【청구항 14】

제9항에 있어서

상기 (d) 단계는,

상기 '검색 깊이(ef_q)'를 다음 수학적 식 7을 통해 동적으로 조정하는 것인

컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램.

[수학적 식 7]

$$ef_q = ef_{low} + \left(\frac{\rho_q - \rho_{min}}{\rho_{max} - \rho_{min}} \right) \times (ef_{high} - ef_{low})$$

여기서, ef_{low} 는 ef_q 의 하한 값, ef_{high} 는 ef_q 의 상한 값, ρ_{min} 는 데이터 밀도의 최솟값, ρ_{max} 는 데이터 밀도의 최댓값을 의미함.

【청구항 15】

제9항에 있어서,

상기 근사 최근접 이웃 검색을 위한 그래프 구축 방법은

상기 데이터셋 내에 모든 데이터 포인트에 대해 상기 (d) 단계 및 상기 (e) 단계를 반복하여 수행하는 단계;를 더 포함하는

컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램.

【청구항 16】

제9항에 있어서

상기 그래프는

HNSW(Hierarchical Navigable Small World) 그래프인 것인

컴퓨터가 판독 가능한 기록 매체에 저장되어 있는 컴퓨터 프로그램.

【청구항 17】

컴퓨터가 판독 가능한 기록 매체에 있어서,

제9항 내지 제16항 중 어느 하나의 항에 따른 컴퓨터 프로그램이 기록된 컴퓨터가 판독 가능한 기록 매체.

【요약서】

【요약】

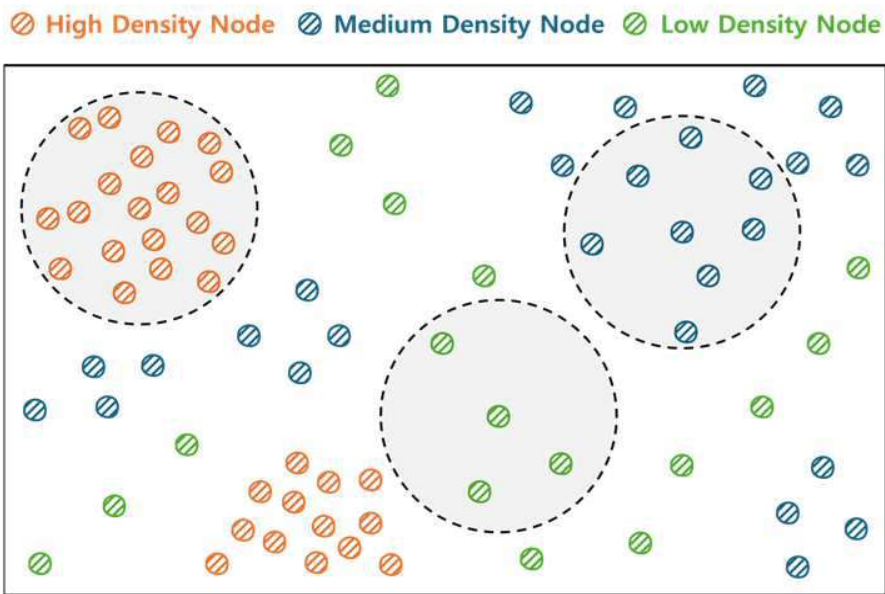
본 발명의 일실시예는 데이터 밀도와 차원 특성에 따라 동적으로 매개변수를 조정하여 근사 최근접 이웃 검색을 위한 그래프를 구축하는 방법에 있어서, (a) 다차원 벡터로 표현된 다수의 데이터 포인트로 구성된 데이터셋 및 초기 매개변수를 수신하는 수신 단계; (b) 수신한 상기 데이터셋의 각 데이터 포인트의 밀도를 계산하는 단계; (c) 상기 밀도의 변동성을 고려하여 매개변수의 범위를 결정하는 단계; (d) 결정된 상기 매개변수의 범위에 기초하여 특정 데이터 포인트의 밀도에 따라 동적으로 상기 매개변수를 조정하는 단계; 및 (e) 상기 특정 데이터 포인트를 조정된 상기 매개변수를 사용하여 그래프에 삽입하는 동적 그래프 구축 단계; 를 포함하고, 상기 매개변수는 '최대 연결 수(M_q)' 및 '검색 깊이(ef_q)'를 포함하는 것인 근사 최근접 이웃 검색을 위한 그래프 구축 방법을 제공한다.

【대표도】

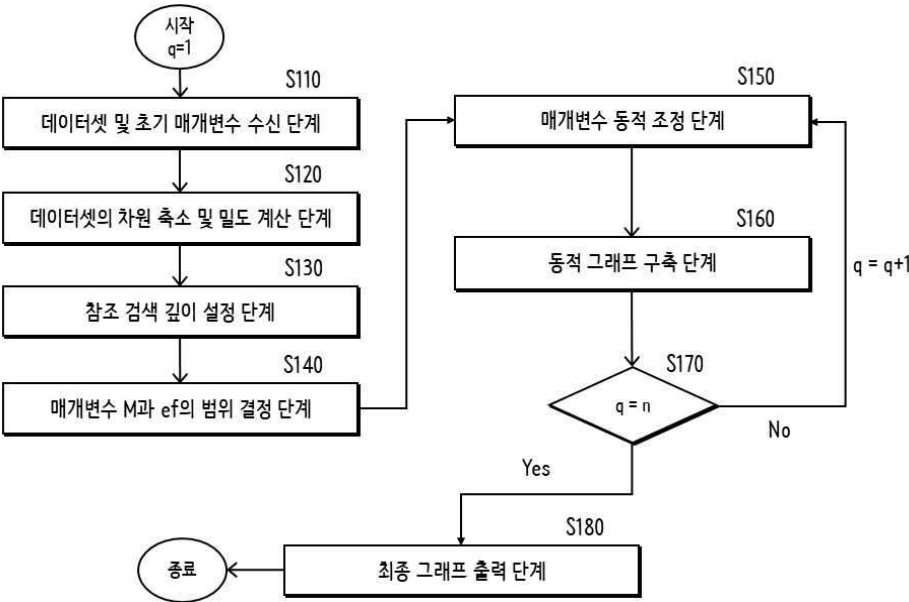
도 2

【도면】

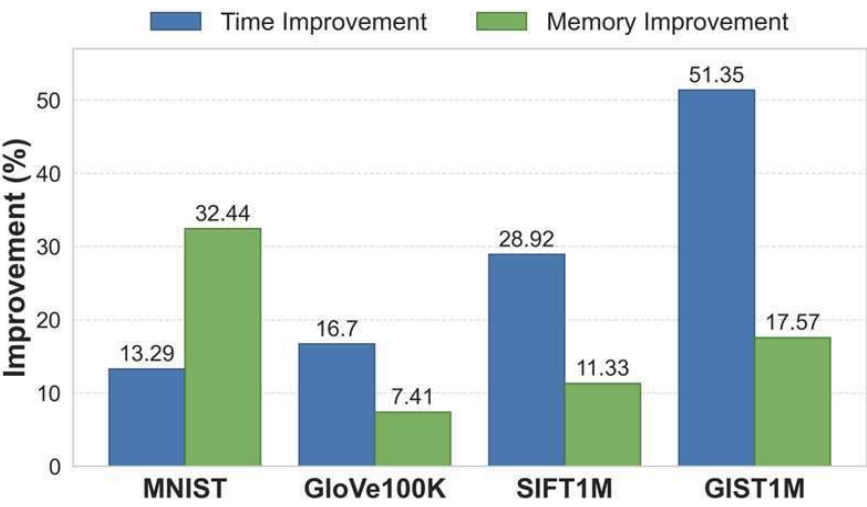
【도 1】



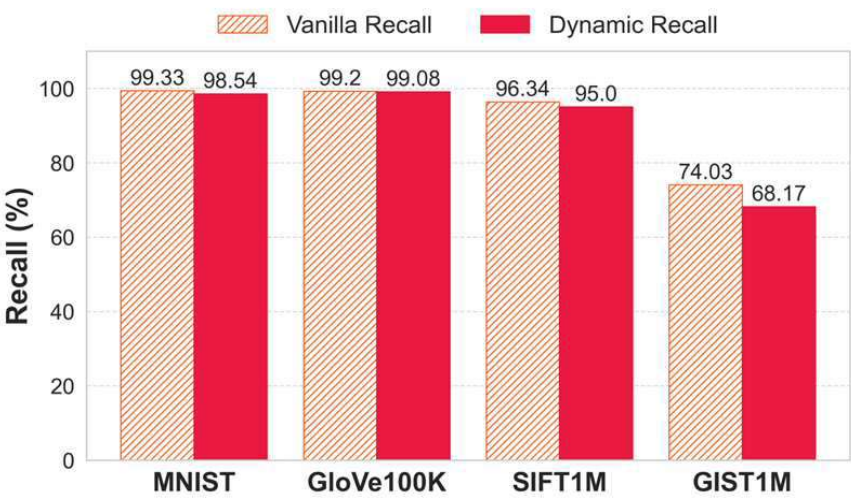
【도 2】



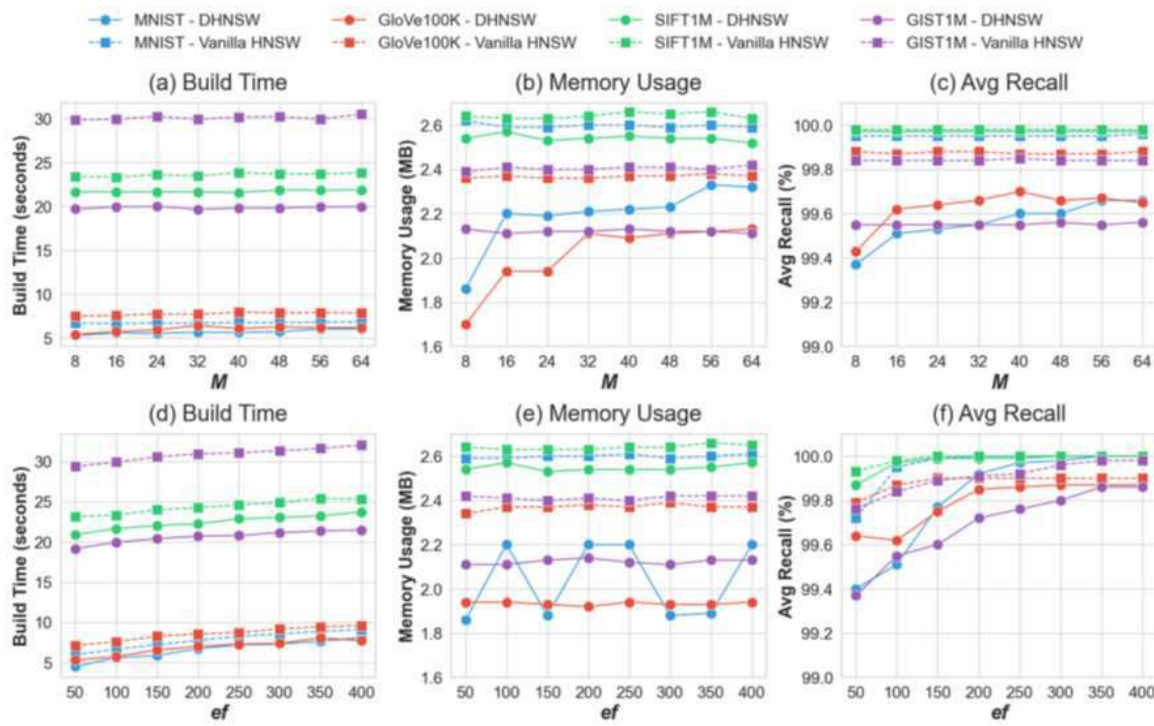
【도 3】



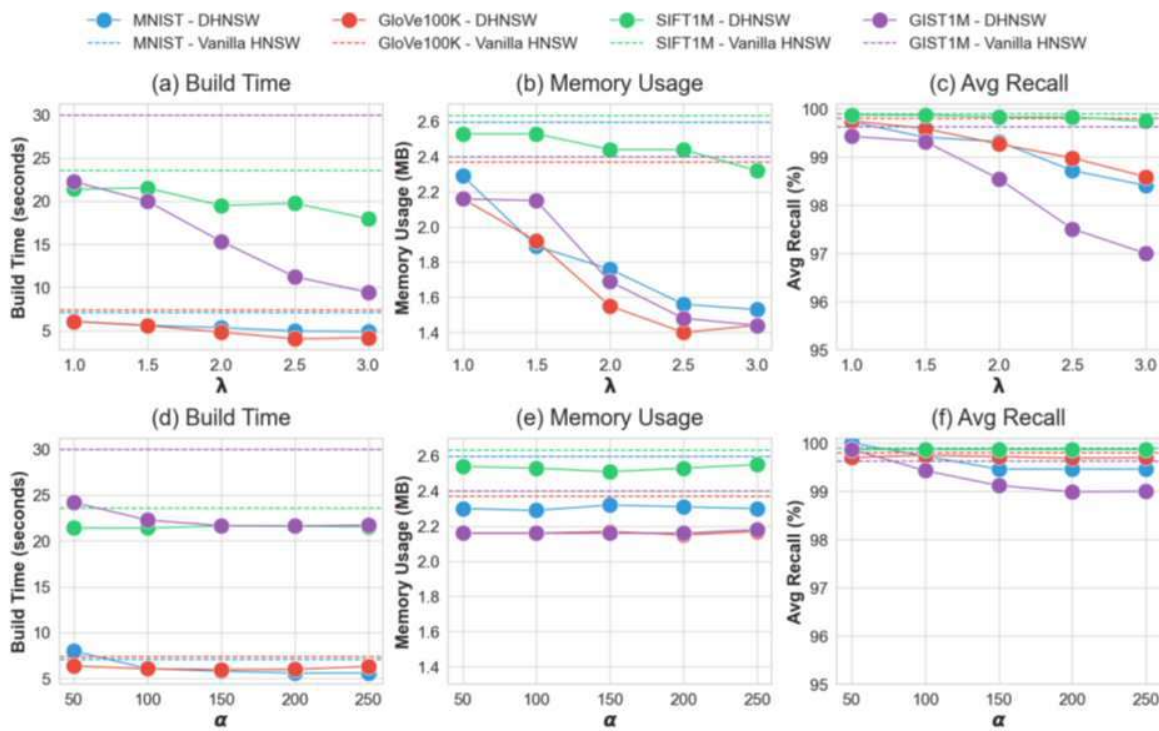
【도 4】



【도 5】



【도 6】



【도 7】

