

FPGA 의 Partial Reconfiguration 기반 동적 밀도 기반 클러스터링 시스템 구현

여리은, 김승희, 김채린, 박유림, *이성주

세종대학교 반도체시스템공학과 및 메타버스융합전공, 세종대학교 전자정보통신공학과, 세종대학교 전자정보통신공학과, 세종대학교 전자정보통신공학과, *세종대학교 AI 융합전자공학과 및 지능형드론융합전공

rieun@itsoc.sejong.ac.kr, seunghee@itsoc.sejong.ac.kr, chaerin1803@sju.ac.kr, yu1578@sju.ac.kr, *seongjoo@sejong.ac.kr

Implementation of Dynamic Density-Based Clustering System using FPGA Partial Reconfiguration

Rieun Yeo, Seunghee Kim, Chaerin Kim, Yurim Park, *Seongjoo Lee

Dept. of Semiconductor Systems Engineering and Convergence Engineering for

Metaverse Sejong Univ., Dept. of Electrical Engineering Sejong Univ. Dept. of

Electrical Engineering Sejong Univ. Dept. of Electrical Engineering Sejong Univ.,

*Dept. of AI Convergence Electronic Engineering and Convergence Engineering for Intelligent Drone, Sejong Univ.

요 약

본 연구는 FPGA 의 Partial Reconfiguration 기능을 활용하여 DBSCAN 과 HDBSCAN 알고리즘 및 KD-Tree 와 Ball-Tree 자료구조를 런타임 중 동적으로 선택 가능한 시스템을 제안한다. 이를 통해 데이터 특성에 따른 최적화된 클러스터링 성능과 FPGA 자원의 효율적인 활용을 실현하였다.

I. 서 론

밀도 기반 클러스터링 알고리즘인 DBSCAN 과 HDBSCAN 은 데이터의 밀도 차이를 고려하여 다양한 형태의 군집을 효과적으로 탐지한다. 최근 연구에서는 KD-Tree 와 Ball-Tree 자료구조를 적용하여 DBSCAN 및 HDBSCAN 의 성능을 향상시켰다. 본 연구는 FPGA 의 Partial Reconfiguration 기능을 활용해 런타임에 알고리즘과 자료구조를 동적으로 전환할 수 있는 시스템을 구현함으로써, 데이터 특성과 차원에 따라 최적의 조합을 선택하여 클러스터링 성능과 효율성을 높이는 것을 목표로 한다.

II. 알고리즘 개요

2.1 DBSCAN 알고리즘 개요

DBSCAN(Density-Based Spatial Clustering of Applications with Noise)은 데이터의 밀도를 기준으로 클러스터를 형성하는 알고리즘이다. DBSCAN 은 두 개의 주요 파라미터, 즉 ϵ (반경)와 \minPts (최소 이웃 포인트 수)를 기반으로 군집을 구성한다. 데이터 포인트가 ϵ 반경 내에 \minPts 이상의 포인트가 존재할 경우 핵심(core) 포인트로 간주하며, 이를 기반으로 밀집된 군집을 형성한다. 반면 밀도 기준을 충족하지 못한 데이터는 노이즈로 구분된다. DBSCAN 은 클러스터의

형태와 크기에 제한을 두지 않아 다양한 형태의 군집을 효과적으로 탐지할 수 있다.

2.2 HDBSCAN 알고리즘 개요

HDBSCAN(Hierarchical DBSCAN)은 DBSCAN 알고리즘의 계층적 확장으로서 다양한 밀도 수준에서 클러스터를 형성한다. HDBSCAN 은 데이터 집합을 계층적 구조로 변환하여 밀도 기반으로 클러스터를 생성하고, 각 클러스터의 안정성을 평가하여 가장 적절한 클러스터를 자동으로 선택한다. 이를 통해 사용자가 별도의 파라미터 설정 없이 데이터에 내재된 다양한 밀도 특성을 반영한 최적의 군집화를 수행할 수 있도록 지원한다. HDBSCAN 은 특히 밀도의 변화가 큰 데이터 환경에서 뛰어난 성능을 보인다.

2.3 KD-Tree 및 Ball-Tree 구조

KD-Tree 는 축 기반으로 데이터를 나누어 빠르게 탐색하는 자료구조로 저차원 데이터에 효율적이다. Ball-Tree 는 공간을 구 형태로 나누어 고차원 및 비정규 데이터에서 효과적으로 작동한다. 이 두 구조는 데이터의 차원과 분포에 따라 선택적으로 사용된다.

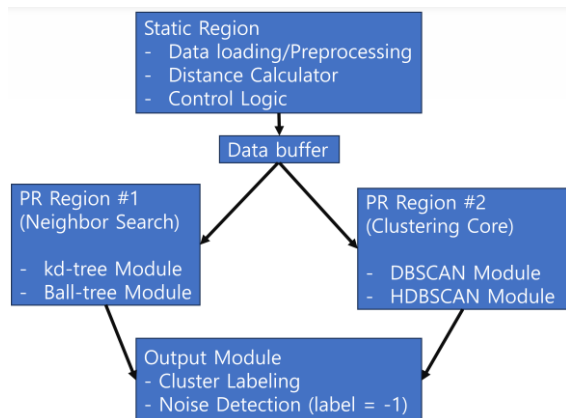
2.4 FPGA Partial Reconfiguration

본 시스템은 FPGA 의 부분 재구성 기능을 사용하여 위의 클러스터링 알고리즘과 자료구조를 런타임에 교체할 수 있도록 설계하였다. 정적인 부분(Static

Region)은 데이터 전처리 및 인터페이스 관리 기능을 담당하며, 가변적인 부분(PR Region)은 알고리즘과 자료구조를 상황에 따라 선택적으로 교체하여 효율성을 극대화한다.

III. 시스템 구현

3.1 시스템 설계 및 구현



본 프로젝트는 밀도 기반 클러스터링 알고리즘(DBSCAN, HDBSCAN)을 FPGA 에서 고속 처리하도록 HLS 로 구현하는 것을 목표로 한다. 입력 데이터는 `points_in[MAX_POINTS][3]`로 주어지며, 출력 데이터는 `labels_out[MAX_POINTS]` 형태이다. 사용자는 `eps`, `min_samples`, `min_cluster_size`, `use_kdtree`, `use_hdbscan` 과 같은 파라미터를 설정할 수 있다.

본 시스템은 Reconfigurable 구조를 기반으로 kd-tree 또는 ball-tree 탐색 방법과 DBSCAN 또는 HDBSCAN 알고리즘을 선택할 수 있도록 설계하였다.

3.2 코드 구조 및 주요 최적화

본 연구에서는 탐색 알고리즘의 내부 코드 중복 제거 및 재활용을 통해 유지보수성과 코드 가독성을 높였다. 거리 계산 루프는 `#pragma HLS UNROLL` 를 이용하여 병렬화하였으며, RAM 의존적인 데이터 구조(queue, labels, neighbors 등)는 `ARRAY_PARTITION` 및 `bind_storage` 를 활용하여 로컬 BRAM 으로 분산하였다. 로직을 `load_points`, `neighbor_search`, `dbscan_cluster` 등으로 나누고 `#pragma HLS DATAFLOW` 를 적용하여 병렬 처리를 가능하게 하였다. 또한, `UNROLL`, `PIPELINE`, `DATAFLOW`, `DEPENDENCE pragma` 등을 적극적으로 활용하여 합성 시간은 증가했지만 성능을 개선하였다.

3.3 Reconfigurable 설계 구성

본 시스템은 FPGA 의 Partial Reconfiguration(부분 재구성) 기능을 활용하여, 클러스터링 알고리즘과 이웃 탐색 자료구조를 런타임 중 상황에 맞게 동적으로 교체할 수 있도록 설계되었다. 전체 구조는 정적 영역과 재구성 영역으로 나뉘며, 정적 영역은 데이터 전처리, 거리 계산기, 인터페이스 제어 등 공통 기능을 고정적으로 수행하고, 재구성 영역은 클러스터링 알고리즘과 이웃 탐색 구조를 선택적으로 탑재하여 유연성을 확보한다.

각 조합은 데이터 특성에 따라 아래 표와 같이 적용된다. 이를 통해 시스템은 데이터 특성에 따라 최적의 연산 조합을 선택하여 자원 효율성과 연산 성능을 극대화할 수 있으며, 전체 Bitstream 을

재로딩하지 않고도 알고리즘을 빠르게 전환할 수 있다는 장점을 가진다.

Kd-tree + DBSCAN	저차원, 균일한 밀도 분포	빠른 군집화 성능, 구조 단순함
Ball-tree + DBSCAN	고차원, 불규칙한 분포에서 고정된 eps 사용 시	고차원 거리 탐색 효율적, 비정규 데이터에 강함
Kd-tree + HDBSCAN	저차원, 다양한 밀도 클러스터를 포함하는 복잡한 데이터	계층적 클러스터링 가능, 안정성 평가 포함
Ball-tree + HDBSCAN	고차원, 밀도 변화가 큰 데이터	다양한 밀도 분포 탐지 가능, 가장 유연함

IV. 결론

제안된 시스템은 FPGA 의 부분 재구성을 이용하여 동적으로 클러스터링 알고리즘 및 탐색 방법을 전환할 수 있도록 구현하였다. 이를 통해 FPGA 의 자원 소비량을 효율적으로 줄이는 데 성공하였으나, HDBSCAN 알고리즘의 하드웨어 리소스 요구량이 매우 크기 때문에 일부 알고리즘의 간소화가 불가피했다. 따라서 향후 연구에서는 FPGA 에 적합한 최적화된 형태의 HDBSCAN 알고리즘 구현에 대한 추가적인 연구가 필요하다. 또한, 현재 시스템은 데이터에 따라 최적의 알고리즘 및 자료구조 조합을 선택할 수 있는 명확한 기준이 설정되지 않았다. 따라서 향후 연구에서는 데이터 특성을 기반으로 최적의 조합을 자동으로 결정하는 기준과 그에 따른 선택 알고리즘을 주제로 심도 있는 연구를 진행할 예정이다.

이를 통해 본 시스템이 더욱 효율적이고 지능적으로 동작할 수 있도록 개선할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

이 논문은 2025 년도 정부 재원(과학기술정보통신부

여대학원생 공학연구팀제 지원사업으로 과학기술정보통신부와

한국여성과학기술인육성재단의 지원을 받아 수행되었습니다.

참 고 문 헌

- [1] T. P. Shibla and K. B. Shibu Kumar, "Improving efficiency of DBSCAN by parallelizing kd-tree using Spark," in *Proc. 2nd Int. Conf. on Intelligent Computing and Control Systems (ICICCS)*, pp. 1197– 1203, Madurai, India, June 2018.
- [2] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825– 2830, Feb. 2011.
- [3] L. McInnes and J. Healy, "Accelerated hierarchical density based clustering," in *Proc. IEEE Int. Conf. on Data Mining Workshops (ICDMW)*, pp. 33– 42, New Orleans, USA, Nov. 2017.