

초병렬 심층 학습 방법 및 장치

(기술분류-인공지능-첨단 시모델링·의사결정(인지·판단·추론))

기술성 분석

기술 개요

- 본 기술은 범용 컴퓨터 클러스터 환경에서 대규모 심층 신경망(Deep Neural Network, DNN) 을 구축하고 고속으로 학습시킬 수 있는 초병렬 심층 학습 방법 및 장치에 관한 것으로, 모델 및 데이터 병렬화를 수행하면서도 delayed gradient 문제를 해결할 수 있음

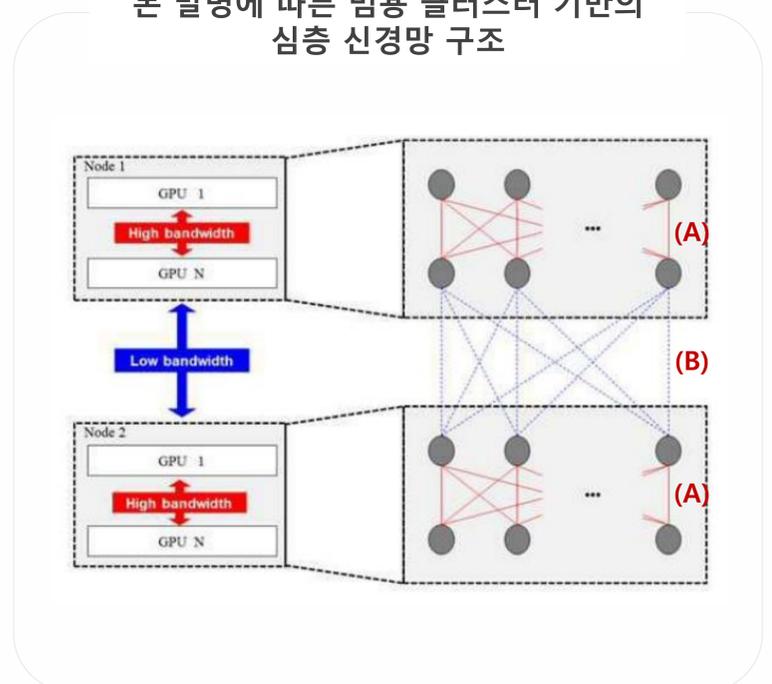
미해결 과제(Unmet needs)

- 대규모 심층 신경망 구축 방법의 필요성
 - 심층 신경망은 뉴런을 층 단위로 쌓아 올려서 인간의 두뇌 구조를 모사한 계산 모델로, 일반적으로 뉴런의 수가 많을수록 높은 성능을 보이나, 뉴런의 수가 증가함에 따라 요구되는 연산량도 증가되므로 단일 서버가 아닌 다수의 서버를 이용한 분산 처리 방식의 병렬 학습이 필요함
 - 심층 신경망의 복잡도는 인간 수준의 지능을 가지기에는 아직도 턱없이 부족하다고 할 수 있으며 강 인공지능과 같은 진정한 인공 지능으로 도약하기 위해서 이제까지 시도되지 않은 대규모 심층 신경망 구조 및 고속의 학습 방법에 대한 연구가 필요함
 - 기존 심층 신경망 및 초대규모 심층 신경망 모델들의 문제점을 회피하면서 대규모 심층 신경망을 구축하기 위해서는 범용 하드웨어를 이용하여 확장성을 가지는 대규모 심층 신경망 구축 방법에 관한 연구가 필요하나, 일반적인 범용 컴퓨터 클러스터를 이용해 신경망의 학습 속도를 개선하는 데에는 어려움이 많음

기술적 해결수단(발명의 구성)

- 1) 본 기술에 따른 범용 클러스터 기반의 심층 신경망 구조
 - GPU가 장착된 컴퓨터 수십 대를 기가비트 이더넷으로 연결하여 범용 클러스터를 구축함
 - (A) 붉은색 실선으로 표시된 GPU 간의 연결은 높은 대역폭(high bandwidth)을 지니는 고속 연결이고, (B) 푸른색 점선으로 표시된 노드 간의 연결은 상대적으로 낮은 대역폭(low bandwidth)을 지니는 연결임
 - 각 컴퓨팅 노드에서는 순전파를 통해 입력 데이터, 파라미터와 할당된 서로 다른 모델을 이용하여 계산한 출력값을 다음 컴퓨팅 노드로 전달하며, 이때 입력 데이터 및 파라미터의 사본을 오류 없는 역전파 계산을 위해 유지함
 - 각 컴퓨팅 노드는 역전파를 통해 입력된 로스(loss), 입력 데이터 및 파라미터를 이용하여 계산한 로스를 다음 컴퓨팅 노드로 전달함

본 발명에 따른 범용 클러스터 기반의 심층 신경망 구조

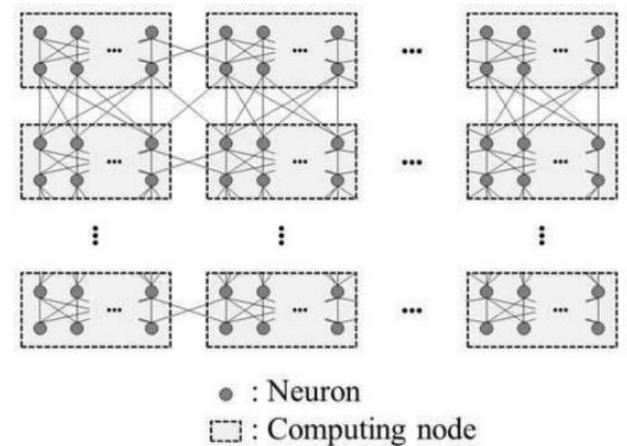


본 기술의 우수성 및 파급 효과

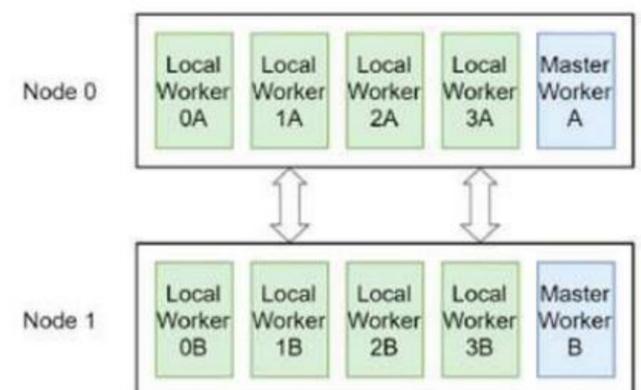
본 기술의 우수성(효과)

- 모델 및 데이터 병렬화 수행
 - 본 기술에 따르면 복수의 컴퓨팅 노드 각각에 연산 수행을 위한 서로 다른 모델을 할당하는 모델 병렬화와, 하나의 컴퓨팅 노드 내에 복수의 로컬 워커(Local Worker)를 배치하여 데이터 병렬화를 수행함
- delayed gradient 문제 해결
 - 컴퓨팅 노드들이 파라미터 서버에게 gradient를 비동기적으로 전송하기 때문에 먼저 도착한 gradient 순으로 마스터 파라미터에 적용되고, 늦게 도착한 gradient는 그 gradient를 계산하기 위해서 사용된 마스터 파라미터가 아니라 먼저 도착한 gradient에 의해서 이미 갱신된 마스터 파라미터에 적용되는 문제, 즉 delayed gradient가 발생함
 - 본 기술에서는 delayed gradient 문제 해결을 위해 Pipelined EASGD(Elastic Averaging Stochastic Gradient Descent) 또는 Pipelined DC-ASGD(Delay Compensation-Asynchronous SGD)를 적용함
 - Pipelined EASGD 기법에서는 각 로컬 파라미터가 마스터 파라미터와 탄성력으로 연결된 것처럼 유지되어 통신 주기마다 동기화되며, delayed gradient의 시간 지연만큼 모델 사본을 만들어 로컬 파라미터가 올바르게 갱신됨
 - Pipelined DC-ASGD는 입력된 로스를 이용하여 gradient를 계산하고 Taylor expansion을 통해 지연되지 않은 시간대의 gradient를 예측하여 최신의 파라미터를 갱신함

모델 병렬화 구조



본 기술에 따른 Pipelined EASGD 구조



적용 제품 및 파급 효과

- 심층 신경망
- 본 기술은 범용 컴퓨터 클러스터 환경에서 대규모 심층 신경망을 구축하고, delayed gradient로 인한 학습 성능 저하 문제를 해결한 초병렬 심층 학습 알고리즘을 제공할 수 있음

지식재산권 현황

발명의 명칭	출원/등록번호	출원/등록일자
초병렬 심층 학습 방법 및 장치	10-2463147	2022.11.01.
패밀리 특허 현황	패밀리 국가	