**제목: EdgePipe: Tailoring Pipeline Parallelism With Deep Neural Networks for Volatile Wireless Edge Devices**

**II. RELATED WORK**

이 논문은 연합 학습(federated learning)을 기반으로 하며, 무선 손실 네트워크의 엣지 디바이스 문제를 다룸. 리소스와 통신 불안정성으로 인해 엣지에서 학습이 느린 문제를 해결하기 위해 파이프라인 가속을 적용함

**A. Federated Learning (연합 학습)**

데이터 병렬성(data parallelism)을 사용🡪 학습 데이터를 여러 하위 작업으로 분할함

데이터 병렬성: 각 장치가 전체 네트워크 모델의 복사본을 가지며 일부 학습 데이터(하위 집합 등)를 학습하는 방식

데이터 병렬성은 신경망이 커질수록 통신 오버헤드가 증가

-> 학습 장치 간의 동기화 문제 발생

엣지 디바이스에서 대규모 모델의 실행 가능성 문제를 해결하기 위해 모델 병렬성(model parallelism)이 연구됨

모델 병렬성: 학습 모델을 장치 간에 분할

-> 리소스 제약이 있는 엣지 디바이스에서도 대규모 학습이 가능

-> 각 장치는 순차적으로 처리 …. 한계: 학습 속도가 느림

이를 가속화하기 위해 Layerwise Staleness, DSP 등의 기법이 도입됨

최근: 하이브리드 병렬성(hybrid parallelism) 접근 (모델 병렬성+데이터 병렬성)

**B. Pipeline Acceleration**

파이프라인 개념을 도입 -> 학습을 가속화하는 연구가 진행됨

PipeDream: 파이프라인 프레임워크를 딥러닝에 적용(딥 러닝 패스로 전환)

장치 간에 순차적으로 전달되는 방식(모델 병렬성처럼..?)으로 처리 성능을 최적화함

GPipe: 미니배치를 마이크로배치로 나누어 학습 시간을 줄임 -> 더 빠르게!!

HetPipe: GPU를 가상 작업자로 나눈다 … 모델과 데이터 병렬성을 결합

XPipe: 데이터 배치를 더 세분화하여 처리함(미니배치는 마이크로배치로 분할)

이러한 병렬화 기법은 비교적 안정적인 네트워크 연결을 전제로 함

무선 엣지 디바이스에서의 불안정한 연결 문제는 지금까지 제대로 다뤄지지 않음

* 무선 네트워크의 변동성에 맞춰 파이프라인 가속을 적용하고, 슈퍼 뉴런(super neuron) 개념을 바탕으로 통신 실패 문제를 해결한 분산 학습 접근법을 제시함

**III. SYSTEM MODEL**

무선 연결을 사용하는 여러 엣지 디바이스를 통한 분산 딥러닝(distributed deep learning) 방식 제시함

손실이 발생하는 네트워크 환경: 불안정한 신경망 관리와 P2P(peer to peer) 통신 오버헤드가 문제

모델 병렬성에서의 자원 미활용 문제 + 엣지 디바이스 특성상 발생할 수 있는 연결 손실 문제

=> 파이프라인 스케줄링으로 해결

파이프라인을 가능한 가득 채움 -> 하드웨어 효율성을 높임

동적 손실 네트워크에서도 학습 모델을 빠르게 구축할 수 있도록 설계

두 가지 핵심 질문을 다룸:

1. 불안정한 네트워크 연결성을 고려한 엣지 디바이스의 이상적인 분할 기준?
2. 디바이스에 분할된 신경망의 훈련을 어떻게 병렬화해 빠른 학습이나 높은 처리량을 얻을 수 있나?

엣지 디바이스는 IEEE 802.11 또는 802.15.4와 같은 저전력 무선 링크를 통해 연결된 것으로 가정함 + 중앙화된 클라우드 액세스는 사용하지 않음

모델 병렬성 기반 -> 모든 데이터셋은 독립적이고 동일한 분포(i.i.d)를 가진 것으로 가정

여러 자원의 계산을 겹쳐서 학습 시간을 단축하고자 함

엣지 네트워크 관점에서 데이터 전송 실패를 완화하도록 모델 설계됨

텍스트, 스크린샷, 도표, 폰트이(가) 표시된 사진

자동 생성된 설명

그림 1. 휘발성 무선 네트워크를 통해 에지 DNN에 대한 파이프라인 병렬성을 활용하는 EdgePipe의 시스템 개요

1. Neural Network Partitioning (신경망 분할):

신경망을 무선 네트워크에 맞게 분할 -> 네트워크가 불안정할 때도 학습이 중단되지 않도록 함: 즉, 네트워크 연결이 불안정해도 모델이 학습을 계속할 수 있도록 탄력적인 구조로 분할..

1. Neuron-to-Device Allocation (뉴런-디바이스 할당):

뉴런을 가장 적합한 디바이스에 매핑: 'super neuron'은 학습에 중요한 뉴런을 의미. 이 뉴런이 **가장 성능이 좋은 디바이스에** 할당됨.

1. Distributed Learning (분산 학습):

미니 배치(데이터 묶음)를 파이프라인 방식으로 처리-> 학습 속도 향상

각 디바이스가 서로 협력 + 데이터 빠르게 처리하는 방식

결론 … EdgePipe: 불안정한 무선 환경에서도 딥러닝 모델이 효율적으로 학습될 수 있도록 **분산 학습, 최적의 디바이스 할당, 신경망 분할** 등을 활용하는 시스템

EdgePipe: 엣지에서 모델 병렬성과 파이프라인 병렬성을 결합-> 확장 가능하고 불안정한 환경에서 학습이 가능하도록 설계됨  
EdgePipe는 두 가지 절차로 구성됨: 모델 병렬성, 파이프라인을 이용한 분산 학습

1. **모델 병렬성**  
   DNN을 부분 서브네트워크로 분할함  
   적절한 계층 간 뉴런들을 각 디바이스에 할당-> 순수한 수직 또는 수평 할당을 피함 + 하나의 레이어 뉴런들이 하나의 디바이스에서 처리되는 것을 방지  
   => 학습 중 통신 오버헤드를 줄임  
   부분 서브네트워크는 엣지 디바이스에 분산 -> 각 디바이스는 자신과 가장 잘 맞는 서브네트워크를 찾음  
   초기 설정 단계에서는 임시 장치 코디네이터로서 최고의 중심성을 가진 노드를 선택함  
   선택된 장치 코디네이터는 신경망 구조를 프로파일링하고 슈퍼 뉴런 기반의 분할 기준에 따라 모델을 분할

**2) 파이프라인 스케줄링을 이용한 분산 학습**  
뉴런이 엣지 디바이스에 할당된 후, 파이프라인 스케줄링을 사용하여 순전파 및 역전파를 처리함  
엣지 디바이스는 이전 레이어의 뉴런 그룹을 담당하는 인접 장치로부터 부분 계산 결과를 받음

**->** 순전파 계산 또는 역전파에서의 기울기 계산을 계속 진행

**->** 이를 다음 레이어를 담당하는 인접 장치에 전송함

결론: **EdgePipe**는 무선 네트워크의 불안정성을 극복

중앙 집중식 클라우드 없이도 에지 장치 간의 효율적인 학습을 가능하게 함