**논문심사의견서(B)에 대한 답변서**

**논문 접수번호:** CST21-11-25-10

**논문 제목:** 훈련 및 검증 성능 개선을 위한 텐서플로우 병렬 처리 기법

심사위원님, 바쁘신 와중에도 본 논문을 검토하고 심사해주셔서 감사합니다.

보내주신 심사의견에 대한 답변은 아래와 같습니다.

국영문 제목이 내용과 부합되고 적당한가? 적당

(지침사항없음)

국영문 초록은 적당한가? 적당

(지침사항없음)

용어의 사용은 올바른가? (표준영어, 번역어의 사용이 정확하고 균일성 여부) 올바름

(지침사항없음)

1. **논문의 구성과 서술 방법이 적당한가? 적당**

(지침사항없음)

연구의 방법과 결과가 분명히 서술되었는가? 잘되었다

(지침사항없음)

과거의 연구와 비교가 되었는가? 수정필요

tf.data에 추가적으로 관련 연구들이 많이 존재할 것으로 생각되므로, 이러한 연구들을 소개하고, 이들과 비교하여 본 논문의 우수성을 보여줄 수 있으면 좋을 것 같음

|  |
| --- |
| **답변** 본 연구와 유사한 관련 연구를 아래와 같이 추가하였습니다.  **[수정사항 반영, 3페이지]** Downpour SGD는 GPU 클러스터 환경에서 훈련데이터와 학습모델의 복제본을 각 GPU에 분배하고 개별적으로 파이프라인을 구성하여 수행하는 기법이다. 갱신된 모든 모델의 가중치는 공유가중치 서버에 전송하여 가중치 평균값을 각 GPU의 모델에 재분배하여 가중치를 반영한다. 이를 통해 모델복사를 사용하지 않는 기존 방식보다 일정 정확도에 도달하는데 60% 빠른 수행시간을 보였다[11]. Nimble은 단일 GPU 환경에서 모델 학습에 필요한 연산을 여러 개의 GPU 스트림(Stream)[12]을 생성하여 분산 처리하는 스케줄링 기법으로, 파이토치에서 기존방식 대비 1.88배만큼 처리속도가 향상되는 결과를 보였다[13].  **[수정사항 반영, 3페이지]** 지금까지의 병렬화 연구들은 학습모델의 훈련 과정을 분산 처리하여 처리속도를 높이는 방법을 제시하였다. 반면, 본 연구에서는 CPU 멀티 쓰레드로 학습모델의 훈련과정과 검증과정을 중첩하여 CPU와 GPU의 활용률을 높임으로써 전체 훈련시간을 개선하는 새로운 기법을 제안한다. |

참고 문헌의 인용이 적합한가? 적당

(지침사항없음)

개발범위(구현 규모, 구현의 중요도)가 충분한가? 불분명

(지침사항없음)

|  |
| --- |
| **답변** 본 연구에서 제안하는 기법의 개발범위를 명확하게 서술하기 위해 아래와 같이 추가하였습니다.  **[수정사항 반영, 7페이지]** 본 제안 기법은 텐서플로우 커널이나 케라스 라이브러리의 기존 코드를 수정하지 않고 사용자 수준에서 Python3.8 버전을 사용하여 구현되었다. 중첩되어 동작하는 훈련 쓰레드와 검증 쓰레드는 Python에서 제공하는 Threading 라이브러리를 사용하였으며 에폭에 대한 훈련 쓰레드가 종료되는 시점에 검증 쓰레드가 동작한다. 검증 쓰레드는 모델복사가 완료되는 시점에 다음 에폭에 대한 훈련 쓰레드를 호출하고 쓰레드풀(Threadpool)을 통해 복사된 모델과 분할된 검증데이터를 분배하여 멀티 쓰레드로 동작한다. 따라서 훈련 쓰레드는 CPU에서 훈련데이터 전처리를 수행하는 동시에 검증 쓰레드는 멀티 쓰레딩을 통해 개별 쓰레드의 작업을 텐서플로우 커널으로 전달한다. 텐서플로우 커널은 요청받은 작업을 여러 개의 GPU 스트림에 할당하여 검증과정을 수행한다[6,12,13]. CPU에서 동작중인 각 쓰레드가 동시에 작업을 요청함에 따라, 이 시점에서 CPU 활용률이 증가하고 검증 과정에서 여러 개의 모델이 병렬적으로 GPU 리소스를 활용함으로써 GPU 활용률 또한 증가하여 전체 수행시간이 단축되는 효과를 보인다. |

논문의 내용이 많은 독자의 관심이 되는가? 보통

(지침사항없음)

서술식 심사평

본 연구는 학습 시 병렬화가 가능한 추론 과정의 전처리 과정을 multi-thread를 이용해서 가속하는 기법을 제안하고, 실제로 이를 구현하여 해당 기법의 효용성을 입증하였습니다. 본 연구에서 보여준 CPU/GPU의 효율 향상 결과는 희망적이나, 본 논문과 유사한 다양한 연구가 존재할 것으로 생각되므로, 관련 연구에 대한 조사 및 비교가 필요할 것 같습니다

|  |
| --- |
| **답변** 본 연구와 유사한 관련 연구를 아래와 같이 추가하였습니다.  **[수정사항 반영, 3페이지]** Downpour SGD는 GPU 클러스터 환경에서 훈련데이터와 학습모델의 복제본을 각 GPU에 분배하고 개별적으로 파이프라인을 구성하여 수행한다. 갱신된 모든 모델의 가중치는 공유가중치 서버에 전송하여 가중치 평균값을 각 GPU의 모델에 재분배하여 가중치를 반영한다. 이를 통해 모델복사를 사용하지 않는 기존 방식보다 일정 정확도에 도달하는데 60% 빠른 수행시간을 보였다[11]. Nimble은 단일 GPU 환경에서 모델 학습에 필요한 연산을 여러 개의 GPU 스트림(Stream)[12]을 생성하여 분산 처리하는 스케줄링 기법을 제안하여 파이토치에서 기존대비 1.88배만큼 처리속도가 향상되는 결과를 보였다[13].  지금까지의 병렬화 연구들은 학습모델의 훈련 과정만 분산 처리하여 처리속도를 높이는 방법을 제시하였다. 반면, 본 연구에서는 CPU 멀티 쓰레드로 학습모델의 훈련과정과 검증과정을 중첩하여 CPU와 GPU의 활용률을 높임으로써 전체 훈련시간을 개선하는 새로운 기법을 제안한다. |