**논문심사의견서(B)에 대한 답변서**

**논문 접수번호:** CPL21-07-05-18

**논문 제목:** 딥 러닝 성능 개선을 위한 쓰레드 기반 셔플링 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? 부적당

본 논문은 딥러닝 성능을 개선하였다고 적시하는데, 어떤 성능을 개선하였는지를 명시하는 것이 필요합니다. 학습 시간과 같이 명확한 성능 향상에 대한 명시가 필요합니다.

|  |
| --- |
| **답변** 개선 부분을 명확하게 하기 위해, 아래와 같이 논문 제목을 수정하였습니다. **[수정사항 반영, 1페이지]** 딥 러닝 훈련 시간 개선을 위한 쓰레드 기반 셔플링 기법  (Thread-based Shuffling Scheme for Improving Training time of deep learning)  또한, 훈련 시간에 대한 명확한 서술을 위해 아래와 같이 초록을 수정하였습니다.  **[수정사항 반영, 1페이지]** 실험을 통해 우리는 제안 기법이 기존 기법 대비 셔플링 시간을 최대 50.7% 그리고 전체 훈련 시간을 최대 13.6% 감소시킬 수 있다는 사실을 확인하였다. 또한, 쓰레드 단위로 데이터를 분할함으로써 발생할 수 있는 정확도 문제를 확인함으로써, 제안 기법이 정확도에 영향을 미치지 않는다는 사실을 확인하였다.    **[수정사항 반영, 1페이지]** Through evaluation, we confirmed that our scheme can reduce the shuffling time by up to 50.7% as well as the total training time by up to 13.6% compared to the conventional approach. Additionally, we also validated the fact that the proposed scheme has no effect on the accuracy issue caused by splitting a range of data into each thread |

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

(지침사항없음)

1. **논문의 구성과 서술 방법이 적당한가? 적당**

(지침사항없음)

연구의 방법과 결과가 분명히 서술되었는가? 수정필요

셔플링은 빅데이터 처리, 메모리 관리 등에서 이미 사용하던 기법입니다. 이에 대한 조사가 추가되어야 하고, 딥러닝 성능 향상에 대한 설명과 비교분석이 추가되어야 합니다. 제안하는 멀티 스레드 데이터 셔플링 기법에 대한 신규성과 우수성, 기여성 등이 추가 설명이 필요합니다.

|  |
| --- |
| **답변** 전체 훈련 시간에 대한 실험을 추가로 진행하고 관련 내용을 그래프와 함께 아래와 같이 추가하였습니다.  **[수정사항 반영, 5페이지]** 마지막으로, 우리는 제안 기법에 대한 전체 훈련 시간에 대한 평가를 위해 추가 실험을 진행하였으며, 그림 7은 실험 결과를 보여준다. 그림 7에서 보여주듯이, 제안 기법이 기본 넘파이 기법 보다 전체 훈련시간을 최대 11.2% 감소시키는 것을 확인할 수 있다. 이는 제안 기법의 셔플링 처리 시간에 대한 단축의 결과로 해석할 수 있다. 그러나, 쓰레드 개수가 증가함에도 불구하고 제안 기법의 전체 훈련 시간이 감소하는 패턴을 보이지 않는다. 우리는 이러한 성능 패턴에 대한 이유를 향후 연구에서 다루고자 한다.    그림 7 훈련 성능 비교  Fig. 7 Comparison of Train Performance |

과거의 연구와 비교가 되었는가? 적당

셔플링은 빅데이터 처리, 메모리 관리 등에서 이미 사용하던 기법입니다. 이에 대한 조사가 추가되어야 하고, 딥러닝 성능 향상에 대한 설명과 비교분석이 추가되어야 합니다.

|  |
| --- |
| **답변** 본 연구와 유사한 관련 연구를 아래와 같이 추가하였습니다.  **[수정사항 반영, 2페이지]** LIRS[4]는 데이터 처리 과정에서 셔플링 오버헤드를 줄이기 위해 SSD에 저장된 훈련데이터를 무작위 순서로 메모리에 할당한다. 그러나 이러한 방식은 저장장치의 I/O을 요청을 빈번하게 요구할 수 있으며, 무작위 순서의 I/O 요청은 저장장치의 성능을 감소시키는 요인 중 하나이다. 이에, 본 논문에서는 딥 러닝 모델 학습의 성능을 개선하기 위한 멀티 셔플링 (Multi-Shuffling) 기법을 어플리케이션 레벨에서 제안한다. |

참고 문헌의 인용이 적합한가? 부족

셔플링은 빅데이터 처리, 메모리 관리 등에서 이미 사용하던 기법입니다. 이에 대한 조사가 추가되어야 하고, 딥러닝 성능 향상에 대한 설명과 비교분석이 추가되어야 합니다.

|  |
| --- |
| **답변** 딥러닝 기법에서 셔플링이 필요한 이유에 대한 내용을 아래와 같이 추가하였습니다.  **[수정사항 반영, 2페이지]** 이는 학습을 위한 훈련 데이터의 고유한 데이터 분포 및 순서 관계로 인해 발생할 수 있는 과적합 (Overfitting) 현상을 방지하기 위함이다.[6][7]  또한, 성능 향상에 대한 설명을 위해 훈련 시간에 대한 실험을 진행하고 관련 내용을 그래프와 함께 추가하였습니다.  **[수정사항 반영, 5페이지]** 마지막으로, 우리는 제안 기법에 대한 전체 훈련 시간에 대한 평가를 위해 추가 실험을 진행하였으며, 그림 7은 실험 결과를 보여준다. 그림 7에서 보여주듯이, 제안 기법이 기본 넘파이 기법 보다 전체 훈련시간을 최대 11.2% 감소시키는 것을 확인할 수 있다. 이는 제안 기법의 셔플링 처리 시간에 대한 단축의 결과로 해석할 수 있다. 그러나, 쓰레드 개수가 증가함에도 불구하고 제안 기법의 전체 훈련 시간이 감소하는 패턴을 보이지 않는다. 우리는 이러한 성능 패턴에 대한 이유를 향후 연구에서 다루고자 한다.    그림 7 훈련 성능 비교  Fig. 7 Comparison of Train Performance |

개발범위(구현 규모, 구현의 중요도)가 충분한가? 불충분

(지침사항없음)

|  |
| --- |
| **답변**  개발 범위 (구현) 관점에서 내용을 보완하기 위해 제안기법 구현과 관련된 내용을 아래와 같이 추가하였습니다.  **[수정사항 반영, 3페이지]** 또한, 쓰레드 생성을 위해 ThreadPoolExecutor API을 사용하였으며, 훈련 데이터의 전체 인덱스를 쓰레드 개수로 분할하여 모든 쓰레드가 동일한 개수의 데이터를 계산하도록 구현하였다. 또한 각 쓰레드 내부에서는 할당 받은 데이터를 넘파이의 shuffle API을 호출함으로써 셔플링 기능을 구현하였다. |

새로운 아이디어의 제시가 있는가? 없다

(지침사항없음)

|  |
| --- |
| **답변** 본 연구에서는 셔플링 시간을 단축함으로써, 전반적인 훈련 시간을 단축시키기 위한 새로운 기법을 제안하였습니다. 그리고 제안 기법의 우수성을 입증하기 위해 추가적인 실험을 진행하였으며, 아래와 같이 관련 내용을 추가하였습니다.  **[수정사항 반영, 5페이지]** 마지막으로, 우리는 제안 기법에 대한 전체 훈련 시간에 대한 평가를 위해 추가 실험을 진행하였으며, 그림 7은 실험 결과를 보여준다. 그림 7에서 보여주듯이, 제안 기법이 기본 넘파이 기법 보다 전체 훈련시간을 최대 11.2% 감소시키는 것을 확인할 수 있다. 이는 제안 기법의 셔플링 처리 시간에 대한 단축의 결과로 해석할 수 있다. 그러나, 쓰레드 개수가 증가함에도 불구하고 제안 기법의 전체 훈련 시간이 감소하는 패턴을 보이지 않는다. 우리는 이러한 성능 패턴에 대한 이유를 향후 연구에서 다루고자 한다.    그림 7 훈련 성능 비교  Fig. 7 Comparison of Train Performance |

논문의 내용이 많은 독자의 관심이 되는가? 보통

(지침사항없음)

서술식 심사평

제목을 보면 기대감이 크나, 본문과 실험결과는 제목과 큰 괴리가 있습니다.  
저자는 내용을 보강하고, 논문의 기여도를 어떤 내용으로 강조할 것인지 고민하여 수정하세요.  
딥러닝 학습, 빅데이터 처리, 데이터 운영의 관점에서 신규성을 찾기 어렵습니다.

|  |
| --- |
| **답변** 개선 부분을 명확하게 하기 위해, 아래와 같이 논문 제목을 수정하였습니다. **[수정사항 반영, 1페이지]** 딥 러닝 훈련 시간 개선을 위한 쓰레드 기반 셔플링 기법  (Thread-based Shuffling Scheme for Improving Training time of deep learning)  또한, 훈련 시간에 대한 명확한 서술을 위해 아래와 같이 초록을 수정하였습니다.  **[수정사항 반영, 1페이지]** 실험을 통해 우리는 제안 기법이 기존 기법 대비 셔플링 시간을 최대 50.7% 그리고 전체 훈련 시간을 최대 13.6% 감소시킬 수 있다는 사실을 확인하였다. 또한, 쓰레드 단위로 데이터를 분할함으로써 발생할 수 있는 정확도 문제를 확인함으로써, 제안 기법이 정확도에 영향을 미치지 않는다는 사실을 확인하였다.  **[수정사항 반영, 1페이지]** Through evaluation, we confirmed that our scheme can reduce the shuffling time by up to 50.7% as well as the total training time by up to 13.6% compared to the conventional approach. Additionally, we also validated the fact that the proposed scheme has no effect on the accuracy issue caused by splitting a range of data into each thread  마지막으로, 신규성에 대한 검증을 위해 추가 실험을 진행하고 아래와 같이 관련 내용을 추가하였습니다.  **[수정사항 반영, 5페이지]** 마지막으로, 우리는 제안 기법에 대한 전체 훈련 시간에 대한 평가를 위해 추가 실험을 진행하였으며, 그림 7은 실험 결과를 보여준다. 그림 7에서 보여주듯이, 제안 기법이 기본 넘파이 기법 보다 전체 훈련시간을 최대 11.2% 감소시키는 것을 확인할 수 있다. 이는 제안 기법의 셔플링 처리 시간에 대한 단축의 결과로 해석할 수 있다. 그러나, 쓰레드 개수가 증가함에도 불구하고 제안 기법의 전체 훈련 시간이 감소하는 패턴을 보이지 않는다. 우리는 이러한 성능 패턴에 대한 이유를 향후 연구에서 다루고자 한다.    그림 7 훈련 성능 비교  Fig. 7 Comparison of Train Performance |