

지능형 엣지 컴퓨팅 시스템을 위한 행렬 연산

손경락, 이정훈
한국의외국어대학교

krson@hufs.ac.kr, tantheta@hufs.ac.kr

Matrix computations for edge intelligence

Kyungrak Son, Jung Hoon Lee
Hankuk University of Foreign Studies

요약

분산 컴퓨팅은 지능형 IoT 네트워크에서 중요한 역할을 하지만 무선 네트워크에서는 제대로 연구되지 않았습니다. 특히 무선 채널과 잡음으로 인한 왜곡은 분산 컴퓨팅의 정확도를 크게 떨어뜨릴 수 있습니다. 안타깝게도 기존의 분산 컴퓨팅 방식은 이 중요한 문제를 적절히 해결하지 못합니다. 본 논문은 지능형 엣지 컴퓨팅 시스템을 위해 연산의 정확도를 개선하는 기법에 대한 논문들에 대해 간략히 요약하여 설명합니다.

I. 본론

무선 환경에서의 고도의 지능형 서비스에 대한 수요는 미래 무선 네트워크 진화의 핵심 동력입니다. 하지만 데이터 분석, 증강/가상 현실(AR/VR), 자율주행, 인공지능을 기반으로 하는 이러한 서비스는 기본적으로 제한된 시간 내에 엄청난 양의 연산을 필요로 합니다. 이처럼 연산량이 많은 무선 서비스를 어디서나 확장 가능한 방식으로 제공하기 위해서는 네트워크의 다른 노드에 연산 작업을 분할하여 할당하고 연산 결과를 취합하는 분산 컴퓨팅이 핵심적인 역할을 할 것으로 예상됩니다.

분산 컴퓨팅의 계산 속도, 정확도 및 계산 부하는 주로 straggler 라고 불리는 가장 느린 노드에 의해 병목 현상이 발생하며, 이러한 노드에 대처하는 것이 중요합니다[1]. 지연 노드에 대처하는 간단하지만 조잡한 방법은 여러 노드에 동일한 작업을 계산하도록 요청한 다음 그중 가장 빠른 노드만 활용하는 것입니다. 그러나 이 방법은 가장 빠른 노드를 제외한 다른 노드는 쓸모가 없기 때문에 원하는 계산 결과를 완벽하게 재구성하는 데 필요한 노드 수 측면에서 매우 비효율적입니다. 이러한 이유로 많은 연구에서 코딩 이론적 접근 방식을 활용하여 특정 수의 노드에서만 받은 결과를 가지고 원래의 대규모 계산 작업을 재구성하는 방법을 사용했습니다 [1]. 현재 코딩 컴퓨팅은 클라우드 네트워크 시스템에서 지연 시간을 줄이고 작업 완료 시간을 개선하기 위해 활용되고 있습니다 [1].

부호화 컴퓨팅은 다양한 유형의 연산과 관련하여 문헌에서 광범위하게 연구되어 왔습니다. 그러나 기존의 다항식 기반 방식은 계산의 분산으로 인한 수치 불안정성 문제가 종종 발생합니다. 이 문제를 해결하기 위해 최근 연구에서는 보간점을 신중하게 선택함으로써 이러한 방식의 수치 안정성을 개선하는 데 초점을 맞추고 있었습니다[2,3,4]. 하지만, [3,4]의 기법들이 잡음이 없는 시스템에서 수치 안정성을 보장하는 연산 방식을 제안하였음에도 불구하고, 잡음이 있는 경우에서는 [5,6]에서 볼 수 있듯이 잡음이 있는 경우 매우 불안정합니다.

계산 속도 측면에서 효율적이라 하더라도 정확도가 떨어지면 그 가치는 상당히 떨어집니다. 이와 관련하여 최근에는 로컬 계산 결과가 노이즈에 의해 손상된 경우에도 행렬 연산의 정확도를 향상시키려는 노력이

이루어지고 있습니다([5,6]). [5]에서는 노이즈 채널의 노이즈를 해결하기 위해 군 대수 및 표현 이론 [7]을 활용하여 보다 강력한 체계를 제안했습니다. [5]에서 각 사용자의 인코딩된 행렬은 동일한 그룹 대수의 함수입니다. 또한 [6]에서는 프레임 양자화 이론을 사용하여 참여하는 작업자 노드의 수를 알 수 없는 경우에도 필요한 재구성 정확도를 보았습니다. 그러나 [5]에서는 시뮬레이션만 있고 관련 이론적 분석은 없습니다. 이러한 이유로 최근 [8]의 논문에서는 이론적 분석과 함께 다양한 네트워크 환경에서 연산의 정확도를 개선하는 논문을 제시하였습니다.

참고 문헌

- [1] K. Lee, M. Lam, R. Pedarsani, D. Papailiopoulos, and K. Ramchandran, "Speeding up distributed machine learning using codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1514-1529, Mar. 2017.
- [2] A. B. Das and A. Ramamoorthy, "Coded sparse matrix computation schemes that leverage partial stragglers," *IEEE Trans. Inf. Theory*, vol. 68, no. 6, pp. 4156-4181, June 2022.
- [3] A. Ramamoorthy and L. Tang, "Numerically stable coded matrix computations via circulant and rotation matrix embeddings," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2684-2703, Apr. 2022.
- [4] M. Fahim and V. R. Cadambe, "Numerically stable polynomially coded computing," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2758-2785, Jan. 2021.
- [5] K. Son, A. Ramamoorthy, and W. Choi, "Distributed matrix multiplication using group algebra for on-device edge computing," *IEEE Sig. Proc. Lett.*, vol. 28, pp. 2097-2101, Oct. 2021.
- [6] K. Son and W. Choi, "Distributed matrix multiplication based on frame quantization for straggler mitigation," *IEEE Trans. Sig. Proc.*, vol. 70, pp. 3058-3073, June 2022.
- [7] H. Cohn and C. Umans, "A group-theoretic approach to fast matrix multiplication," in *Proc. of IEEE Symp. Found. of Comput. Sci. (FOCS)*, Cambridge, MA, USA, Oct. 2003, pp. 438-449.
- [8] K. Son and W. Choi, "Coded Matrix Computation in Wireless Network," *IEEE Trans. on Wireless Commun.*, Early access