Recruiting Notice:
Hiring RA, Student Workers, Volunteers, PM for Machine Learning Algorithms and Systems

Dear Students,

We are from Information Theory and Machine Learning (vITAL) research lab and USC-Amazon Center on Secure and Trusted Machine Learning, directed by Professor Salman Avestimehr. Our current mission is to build trustworthy, decentralized, automated, system-efficient ML algorithms and systems for diverse AI applications in computer vision, natural language processing, data mining, IoT, and 5G.

Now we are hiring **paid** RA/student workers/volunteers to join our research group. We have various positions for graduate/undergraduate students.

Position 1: Research on Distributed/Federated Machine Learning Algorithms and System
Position 2: Research on Trustworthy ML (Security, Privacy, Fairness, etc)
Position 3: Research on ML for Diverse AI Applications (CV, NLP, GraphNN, IoT, etc)
Position 4: ML System Design and Optimization (Distributed System/Cloud Computing)
Position 5: ML System Design and Optimization (Mobile Computing)
(For the responsibilities and qualifications of each research position, please read details at this link: https://docs.google.com/document/d/1KoC_tOUdF5Yxzyc_sGndAMx_8NjBKr3ixSQKMxsdXIo/view)

We encourage you to apply, if you are in one of the following situations:

1. You are already a PhD student now but hope to explore research topics similar to our mission. Please apply to work as an RA in our lab.
2. You are an undergraduate or a master student, and want to apply to a PhD program after graduation.
3. You are an undergraduate or a master student, and want to get more experience in engineering and development, which helps you to look for an industrial job such as general software engineer or ML engineer after your graduation.

To apply, you can email Chaoyang He (chaoyang.he@usc.edu) and CC Professor Salman Avestimehr (avestime@usc.edu), with the title “Applying to Work on ML Algorithms and System”. In the email, please (1) attach your latest CV, (2) introduce your goal to work with us (getting experience for publication or industrial jobs), (3) your experience related to the position (course, project, paper, intern, etc.), and (4) how long and when you can devote to our project (e.g, 20 hours/week, Thursday to Friday) and (5) which position do you prefer. We will schedule a meeting with you if we match well.

Position 1: Research on Distributed/Federated Machine Learning Algorithms and System

[Responsibilities]

This is a research-oriented role. You will explore one of the following research directions:

1. Large-scale distributed learning training algorithm and system for massive-scale models (e.g., Transformers, BERT, ViT, MoE, etc)
2. Federated learning algorithms towards resource-constraint, personalization, robustness, weakly-supervision, etc.
3. Neural architecture design towards efficient distributed/federated training/inference with techniques such as Neural Architecture Search (NAS), Model compression with knowledge distillation, quantization, low-rank etc.

Our target conferences are ICML, NeurIPS, ICLR, CVPR, ACL/EMNLP, MLSys, SOSP/OSDI, etc. To have a sense of research taste, please check out our publications:

[1] PipeTransformer (ICML'2021, https://arxiv.org/abs/2102.03161)

[2] FedML (NeurIPS 2020 FL, Best Paper Award, https://arxiv.org/abs/2007.13518)

[3] FedGKT (NeurIPS 2020, https://arxiv.org/abs/2007.14513)

[4] MiLeNAS, FedNAS (CVPR 2020, https://arxiv.org/abs/2003.12238)

[Qualifications]

1. Experience in machine learning optimization, computer vision, natural language processing, data mining, etc
2. Experience in neural architecture design or optimization for Transformers (BERT, ViT), Graph Neural Networks, CNNs, etc
3. Experience in distributed/federated machine learning is a plus, but not required.
4. Experience in AutoML (NAS) and model compression is a plus, but not required.
5. Good programming skills in Python, PyTorch, TensorFlow, etc

Position 2: Research on Trustworthy ML (Security, Privacy, Fairness, etc)

[Responsibilities]

This is a research-oriented role. You will explore one of the following directions:

1. secure aggregation algorithm
2. adversarial attack and defense, robustness for federated learning

To have a flavor of our publications in this track, please read the following papers:

[1] Turbo-Aggregate: https://arxiv.org/abs/2002.04156

[2] A Scalable Approach for Privacy-Preserving Collaborative Machine Learning, accepted to NeurIPS, 2020

[3] Securing Secure Aggregation: Mitigating Multi-Round Privacy Leakage in Federated Learning. https://arxiv.org/abs/2106.03328

[4] Secure aggregation with heterogeneous quantization in federated learning https://arxiv.org/abs/2009.14388
[5] Mitigating Byzantine Attacks in Federated Learning. https://arxiv.org/abs/2010.07541
[6] CodedPrivateML: A Fast and Privacy-Preserving Framework for Distributed Machine Learning. https://ieeexplore.ieee.org/abstract/document/9330572

[Qualifications]

1. Experience in security and privacy, such as adversarial attacks, secure aggregation;
2. Experience in deep learning training frameworks such as PyTorch and TensorFlow;
3. Basic understanding in Transformers, CNNs, Graph Neural Networks (you don't need to design new models)
4. Experience in building distributed computing program (e.g., distributed computing with MPI)
5. It is a plus if you have experience in mathematical proof for security and privacy.

Position 3: Research on ML for Diverse AI Applications (CV, NLP, GraphNN, IoT, etc)
[Responsibilities]
This is a research-oriented role. It has a research flavor of application-driven exploration, but you need to design algorithm and models to improve the performance (accuracy, efficiency, etc). Our target conferences are CVPR/ICCV/ECCV, ACL/EMNLP, KDD/AAAI. To have a sense of research taste, please check out our publications:
[1] SpreadGNN: https://arxiv.org/pdf/2106.02743.pdf
[2] FedNLP: https://fedml.ai/files/FedNLP.pdf
[3] FedCV: https://fedml.ai/files/FedCV.pdf
[4] FedGraphNN: https://arxiv.org/abs/2104.07145
[5] FedIoT: https://arxiv.org/pdf/2106.07976.pdf
(we already finish the library and benchmark. In the next phase, we will upgrade these libraries from the perspective of algorithms and modeling)

[Qualifications]

1. Deep understanding important deep learning models such as Transformers, CNNs, Graph Neural Networks, VAE, GAN, etc
2. Experience in machine learning optimization, especially distributed optimization analysis
3. Good programming skill in PyTorch or TensorFlow
4. Enthusiastic in data analysis
5. Enthusiastic in quantitative analysis to improve the model performance
6. Good insights in identifying application opportunities for various deep learning models and applications.

Position 4: Machine Learning System Design and Optimization (Distributed System/Cloud Computing)
[Responsibilities]
This is a development-orient role. You will build distributed system or cloud computing for large-scale machine learning pipeline, especially for large-scale federated learning, distributed training and inference. You don't need to publish paper but you need to assist researchers to

develop and optimize the system. To have a flavor of the system development, please check these two publications:
[1] PipeTransformer (ICML'2021, https://arxiv.org/abs/2102.03161)
[2] FedML (NeurIPS 2020 FL, Best Paper Award, https://arxiv.org/abs/2007.13518)

[Qualifications]

1. Experience in building distributed system, cloud computing system, etc.
2. Strong programming skills in Java, C++, Python, Javascript, etc. You don't need to know all, but we expect you are good at two of them.
3. Experience in machine learning, CV, NLP is not required, but we hope you have some basic understandings of ML system
4. Experience building maintainable and testable code bases, including API design and unit testing
5. Experience in networking APIs (HTTPs, TCP, UDP, MQTT, MPI etc.); build cloud service with stable open source frameworks.
6. Experience in multithreading/multi-process programming and system resource management

Position 5: Machine Learning System Design and Optimization (Mobile Computing)
[Responsibilities]
In this role, you will be responsible for developing and optimizing machine learning algorithms and SDKs in Android/IoT platform. The output contains both research publication and open source library. Product delivery will be our next phase focus. To understand our vision, please read the paper: https://arxiv.org/abs/2007.13518

[Qualifications]

1. Experience building Android applications in Java using Android SDK
2. Experience in C++ for Android NDK is not required, but we hope you could handle it by fast learning.
3. Experience in machine learning, CV, NLP is not required, but we hope you have some basic understandings of ML system
4. Experience building maintainable and testable code bases, including API design and unit testing
5. Experience in networking APIs (HTTPs, TCP, UDP, MQTT etc.)
6. Experience with multithreading programming and system resource management; Good at using Android *Service* component
7. Experience in Java/C++-based Deep Learning framework is a plus
7. Enthusiastic to system performance optimization