AKSHAY MEHRA

https://akshaymehra24.github.io/ | +1 (614) 815 5329 | amehra@tulane.edu

Education

Tulane University — Aug 2019 – May 2023 (expected)

Ph.D. in Computer Science — **GPA: 3.97**

Advisor: Prof. Jihun Hamm

Research Interests: Adversarial machine learning and robustness, out-of-distribution robustness, bilevel optimization, generalization in deep neural networks.

- Understanding the limitations of learning in the unsupervised domain adaptation (UDA) setting and developing methods to improve their robustness to distributions not conducive for UDA. Recently, we demonstrated that current UDA methods are highly vulnerable to data poisoning attacks, suggesting learning in the UDA setting is extremely challenging. Paper

- Understanding the robustness of machine learning models trained with certified defenses to data poisoning attacks. We proposed a new bilevel-optimization based attack that modifies the training imperceptibly and leads to significant reduction in the robustness guarantees of certifiably robust machine learning models. Paper

- We proposed an efficient method for solving large-scale constrained bilevel optimizations appearing in the field of machine learning. The method has smaller space and time complexity compared to previously proposed methods and can solve bilevel problems involving several million variables easily. We demonstrated good performance on various tasks including data denoising by importance learning, few-shot learning, and training-data poisoning. Paper

- Analyzing the reason for existence of adversarial examples in several machine learning / deep learning models from an optimization / game theoretic standpoint. Proposed an alternative method to alternating minimax using sensitivity term which aims to stabilize the search for the minimax point Paper. Interested in developing theoretical foundations for adversarial robustness and understanding how it is different from traditional generalization.

The Ohio State University — Aug 2016 – Aug 2019

Master of Science in Computer Science — **GPA: 4.0**

Advisors: Prof. Jihun Hamm and Prof. Mikhail Belkin

Used large-scale graph based semi-supervised learning with multimodal features to improve concept retrieval. Augmented semi-supervised learning with active learning to generalize to user's concept using only a handful of labeled points. Demonstrated effective performance on the ImageNet and AnimalWithAttributes dataset. Paper

Relevant Courses: Non-linear Optimization, Artificial Intelligence, Machine Learning, Speech and Language Processing, Probability for Statistical Inference, Deep Learning Applications, Advanced Algorithms, Advanced Operating Systems, Programming Languages, Real Analysis, Probability Theory.

Thapar University — July 2011 – June 2015

Bachelor of Engineering in Computer Engineering — **GPA: 9.71 / 10**

Relevant Courses: Data Structures, Algorithms, Advanced Data Structures, Graph Theory, Artificial Intelligence, Numerical Analysis, Linear Algebra, Optimization, Operating System, Networking, Databases, Object Oriented Programming.

Publications

- Understanding the Limits of Unsupervised Domain Adaptation via Data Poisoning.
  Akshay Mehra, Bhavya Kailkhura, Pin-Yu Chen and Jihun Hamm.
  Neural Information Processing Systems (NeurIPS) 2021.
- How Robust are Randomized Smoothing based Defenses to Data Poisoning?.
  Akshay Mehra, Bhavya Kailkhura, Pin-Yu Chen and Jihun Hamm.
  Computer Vision and Pattern Recognition (CVPR) 2021.
- Penalty Method for Inversion-Free Deep Bilevel Optimization.
  Akshay Mehra and Jihun Hamm.
  Asian Conference on Machine Learning (ACML) 2021.
- Fast Interactive Image Retrieval using large-scale unlabeled data.
  Akshay Mehra, Jihun Hamm and Mikhail Belkin.
- Machine vs Machine: Minimax-Optimal Defense Against Adversarial Examples
  Jihun Hamm and **Akshay Mehra**.

Internships and work experience

Lawrence Livermore National Laboratory

May 2020 – Aug 2020 &
May 2021 – Aug 2021

Research Intern (Supervisor: Dr. Bhavya Kailkhura)
Data poisoning attacks against certified defenses.

- We studied the problem of using data poisoning attacks to affect the robustness guarantees of classifiers trained using certified defense methods.
- We proposed a bilevel-optimization based attack which can generate poison data against several robust training and certification methods. We specifically used the attack to highlight the vulnerability of randomized smoothing based certified defenses to data poisoning.
- We demonstrated the effectiveness of our attack in reducing the certifiable robustness obtained using randomized smoothing on models trained with state-of-the-art certified defenses such as Gaussian data augmentation, SmoothAdv and MACER. For these training methods our attack reduced the average certified radius of the target class by more than 30\%.

Avata Intelligence

May 2019 – Aug 2019

Summer Intern (Supervisors: Dr. Manish Jain & Dr. Matthew Brown)
Train and Test time attacks on Neural Networks:

- Evaluated effectiveness of various test time attacks on fully connected and convolutional neural networks. Compared the performance of these methods against a margin-based attack. Our margin-based attack produces adversarial examples with lesser distortion and is twice as fast when compared to the CW attack.
- Compared performance of data poisoning attacks generated by solving Bilevel Optimization against other methods such as reverse mode automatic differentiation. Our method for solving Bilevel Optimization produces significantly better results using only a about 6% poisoning points.

Microsoft

May 2017 – Aug 2017

Data Scientist Intern (Supervisors: Dr. Mohamed Abdel-Hady & Dr. Debraj GuhaThakurta)
Entity Extraction from Bio-Medical Data, to identify Drugs, Diseases etc. from medical records. Published here.

- Trained Word2Vec model using 15 million Medline Abstracts to extract word embeddings for medical entities using Spark in < 90 mins.

- Build a LSTM based recurrent neural network to train the entity extractor using Keras with Tensorflow backend. RNN initialized with Medline embeddings outperformed the RNN initialized with generic word embeddings trained on Google News. Our model detected 7012 entities correctly (out of 9475) with a F1 score of 0.73 compared to 5274 entities with F1 score of 0.61 obtained with RNN initialized with generic word embeddings. Code here.

Microsoft

June 2015 – Aug 2016

Software Engineer

Secure License Keys (SLK), a project involving generation and distribution of product keys for almost all Microsoft products.

- Optimization of product key inventory using various machine learning techniques. Obtained mean absolute percentage error of 5.84 for the next peak quantity and 1.54 for next peak time. Improved the system by about 25%.
- Utilized telemetry data of several internal applications for Anomaly Detection for proactive monitoring of systems.

Microsoft

June 2014 – July 2014

Software Engineer Intern

Reporting on Cloud, a Hybrid BI Portal, aimed at being a one stop shop for all reporting needs of a business user.

- Developed a SharePoint portal and integrated existing reports with Power View. Showed how effective visualization helps business understand support contract renewal by premier customers of Microsoft. Provided actionable insights for the PFE's to ensure majority contract renewals, using Machine Learning models.

Projects

Relation Extraction for Drug-Drug Interaction

Sept 2017 – Dec 2017

- Modeled Drug-Drug Interactions using Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) based deep learning models augmented with word level Attention to understand how administration of two drugs affect each other.
- Used pre-trained word embeddings along with position embeddings for drug entities, embeddings for Part-of-Speech tags to effectively model the input sentence. Obtained a Macro-Averaged F1 score of 0.73 with a two-stage model which is very close to the state-of-art performance of 0.77

LexComp: Single Document Summarizer

Feb 2017 – Apr 2017

- A novel technique for single document summarization that uses Lex Rank and integer linear programming for sentence compression.
- We obtain a Rouge score of 0.34 using this method on the New York Times dataset which is an improvement over using only Lex Rank

Predicting Influencers in a Social Network

Oct 2016 – Dec 2016

- Achieved an AUC of 0.869 using Gradient Boosting and an AUC of 0.862 using Logistic Regression compared to an AUC on 0.88 of the winning solution on Kaggle' s Competition

Technical Skills

Programming, Databases, Tools: C, C++, Python, Tensorflow, Keras, Shell Scripting, Azure, SQL Server, Oracle, MySQL.

Competitive Programming Handles: Codechef (rihaan1991), Hackerrank (akshaymehra), Google Code Jam (akshay1622).

Awards

1. Recipient of the SLK **team award** under the category "Do Great Work" by Microsoft, 2016.
2. Team selected to represent the university at **ACM-ICPC** at Amritapuri Site, India in Nov 2014.

3. Recipient of **full Tuition Free Ship by Thapar University for all 4 years**, 2011-2015.
4. Runner Up of Code-Uncode: India’s first secure coding competition, EC-Council 2014