Adithya Kulkarni

470.331.0372 | aditkulk@iastate.edu | https://github.com/kulkarniadithya |
https://www.linkedin.com/in/adithya-kulkarni-581943a0/ | Ames, Iowa

EDUCATION

Iowa State University Jan 2020 – May 2024 (expected)
Ph.D in Computer Science Ames, IA, USA

- **Research Area**: True label inference, Data Mining, Machine Learning, Natural Language Processing
- **Advisor:** Dr. Qi Li

Iowa State University Jan 2020 – Dec 2021
M.S in Computer Science Ames, IA, USA

- **Research Area**: True label inference, Data Mining, Machine Learning, Natural Language Processing
- **Advisor:** Dr. Qi Li
- **Thesis:** True label inference from heterogeneous data sources in Natural Language Processing

National Institute of Technology, Agartala Jun 2012 – May 2016
B.Tech in Computer Science and Engineering Agartala, TR, India

RESEARCH INTERESTS and ACHIEVEMENTS

My research interests include weak supervision and true label inference, focusing on data mining, machine learning, and natural language processing. Specifically, my research goal is to obtain high-quality labels for a corpus by reducing data annotation costs. I have contributed several key methods in top NLP and data mining conferences through my research. I have **six published works** in top conferences such as **UAI, ICDM, SDM, ACL, SIGKDD, and EMNLP findings**, and have submitted multiple works to **EMNLP** which are under review.

CrowdSourcing and True Label Inference

Machine Learning models are data-hungry and need large amounts of labeled data for training. Crowdsourcing platforms such as Amazon Mechanical Turk (AMT) are used to cheaply hire non-expert human annotators to annotate unlabeled corpora, thus minimizing the annotation costs. Since crowd workers are non-experts, repeated labeling is done for each sample in the corpus to reduce labeling errors. Repeated labeling combined with large amounts of data can result in high annotation costs. However, some samples in the unlabeled corpora may have label correlations, and this correlation can significantly reduce annotation costs. Obtaining crowd labels for a few correlated samples can help infer labels for the remaining correlated samples. Through my research, I have been working on projects to identify the label correlation between samples in the unlabeled corpora by modeling the correlation as a graph structure and utilizing the computed label correlation to choose important samples to obtain crowd labels and then infer labels for the remaining correlated samples by propagation labeling information. The research projects have resulted in **one publication** in **UAI, 2023**, a highly competitive artificial intelligence conference.

Finding trustworthy information and reliable workers among crowd workers without supervision is challenging. To address this challenge, I have been working on truth discovery and Ising model based projects driven by applications in natural language processing. The projects resulted in **three publications** in highly competitive data mining and NLP conferences, including **ICDM, 2021, SDM, 2022, and EMNLP findings, 2020**. The inherent assumption of crowdsourcing tasks is that the workers are considered independent. This assumption may not hold if generative machine learning models are used as crowd workers. My recent projects focus on testing this assumption and proposing modifications to existing aggregation methods to accommodate generative machine learning models as crowd workers.

Review Analysis

Almost all businesses that offer various services to customers obtain feedback for their services through customer reviews. Through reviews, customers generally provide their opinions and suggestions that can help improve the services rendered. Extracting this important and useful information from unstructured and abundant reviews is pivotal for the success of businesses. Through my projects, I have proposed several unsupervised (zero-shot prompt-based) and weakly supervised methods to label the important information in the reviews. The weak labels obtained using my proposed approaches can have noise and bias. Therefore, I have proposed an approach to tackle the issues of noise and bias in the weak labels so that machine learning models can be properly trained using the generated weak labels. The research projects resulted in **two publications** in highly competitive NLP and data mining conferences, such as **ACL, 2023 and SIGKDD, 2022**.

Fact Verification

In this age of the Internet, several prominent individuals make multiple claims on various subjects to the press or on social media platforms that can have a wide range of effects on the general public. Real-time automatic fact-verifiers are proposed for various platforms that fact-verify these sensitive claims. With the recent success of ChatGPT, I have been working on projects to validate if generative AI models like ChatGPT can be used for sensitive fact-verification tasks. Specifically, I have tested different prompt designs that can enhance the performance of ChatGPT for sensitive fact-verification tasks. The projects resulted in a paper submitted to the **EMNLP, 2023** conference.

TEACHING

Teaching Assistant

Iowa State University Ames, IA, USA

As a teaching assistant, my duties include assignment creation including solutions, exam proctoring, grading, office hours and recitations.

COM S 311 - Undergraduate Design and Analysis of Algorithms *Spring 2020, Fall 2020 and Summer 2021*

- Taught one undergraduate class of 70 students during Summer 2021.

COM S 331 - Undergraduate Theory of Computation *Fall 2021*

- Taught 45 undergraduate students as part of recitation classes.

COM S 531 - Graduate Theory of Computation *Spring 2021 and Fall 2023*

- Taught 20 graduate students as part of supplemental instruction classes.

HONORS AND AWARDS

- **Research Award:** Graduate research excellence award for outstanding research contributions. (December, 2022)
- **Teaching Award:** Graduate teaching excellence award for graduate theory of computation course. (May, 2021)
- **Poster Presentation:** Second place at the 5th Annual Research Day Competition. (May, 2021)
- **Travel Award:** Recipient of SIAM travel awards to attend the conference virtually and present my paper. (April, 2022)

TEACHING CERTIFICATIONS

- Preparing Future Faculty Associate. (May, 2022)
- Preparing Future Faculty Fellow. (May, 2022)
- Preparing Future Faculty Scholar. (May, 2022)

PUBLICATIONS

Conferences

UAI, ACL, SIGKDD, SDM, ICDM, EMNLP-Findings, EMNLP

- Optimal Budget Allocation for Crowdsourcing Labels for Graphs. **Adithya Kulkarni**, Mohna Chakraborty, Sihong Xie, Qi Li, **Thirty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI), 2023**, url: https://proceedings.mlr.press/v216/kulkarni23a.html. (Acceptance Rate: 29.3%)
- Zero-shot Approach to Overcome Perturbation Sensitivity of Prompts. Mohna Chakraborty*, **Adithya Kulkarni***, Qi Li, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL), 2023**, url: https://aclanthology.org/2023.acl-long.313. (Acceptance Rate: 20.8%)
- Open-Domain Aspect-Opinion Co-Mining with Double-Layer Span Extraction. Mohna Chakraborty*, **Adithya Kulkarni***, Qi Li, **SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD), 2022**, url: https://doi.org/10.1145/3534678.3539386. (Acceptance Rate: 14.9%)
- CPTAM: Constituency Parse Tree Aggregation Method. **Adithya Kulkarni***, Nasim Sabetpour*, Alexey Markin, Oliver Eulenstein, Qi Li, **SIAM International Conference on Data Mining (SDM), 2022**, url: https://doi.org/10.1137/1.9781611977172.71. (Acceptance Rate: 21.3%)
- Truth Discovery in Sequence Labels from Crowds. Nasim Sabetpour*, **Adithya Kulkarni***, Sihong Xie, Qi Li, **IEEE International Conference on Data Mining (ICDM), 2021**, url: https://doi.org/10.1109/ICDM51629.2021.00065. (Acceptance Rate: 9.77%)
- OptSLA: An Optimization-Based Approach for Sequential Label Aggregation. Nasim Sabetpour*, **Adithya Kulkarni***, Qi Li, **Findings of Empirical Methods in Natural Language Processing (EMNLP findings), 2020**, url: https://doi.org/10.18653/v1/2020.findings-emnlp.119. (Acceptance Rate: 12.7%)
- An Empirical Study of Using ChatGPT for Fact Verification Task. Mohna Chakraborty*, **Adithya Kulkarni***, Qi Li, **Empirical Methods in Natural Language Processing (EMNLP), 2023** (Under Review)
- Unsupervised Universal Dependency Parse Tree Aggregation Applying Ising Model. **Adithya Kulkarni**, Oliver Eulenstein, Qi Li, **Empirical Methods in Natural Language Processing (EMNLP), 2023** (Under Review)

CONFERENCE PRESENTATIONS

- Optimal Budget Allocation for Crowdsourcing Labels for Graphs, UAI, 2023 (virtual poster presentation)
- CPTAM: Constituency Parse Tree Aggregation Method, SDM, 2023 (virtual conference presentation)

SERVICE

- Student representative for computer science department in Liberal Arts and Science (LAS) graduate student committee (August, 2023 - present).
- Reviewer of proposals for computer-based instructional support as part of Liberal Arts and Science Committee on the Advancement of Student Technology for Learning Enhancement (LASCASTLE) (February, 2022)
- President of Computer Science Graduate Student Organization (CSGSO) (August, 2022 - August, 2023).
- Vice-President of Computer Science Graduate Student Organization (CSGSO) (August, 2021 - August, 2022).
- Reviewers for SDM, KDD, PAKDD, EMNLP, ACL, WSDM, HCOMP conferences.
- Volunteer of Cymath program at Iowa State University aimed at taking extra math classes for underrepresented school students of grades 3 and 4.
- Co-founder of a student empowered non-profit organization **WeCan** at National Institute of Technology (NIT), Agartala. The vision of WeCan is to empower economically backward students in the neighboring villages of NIT Agartala.

WORK EXPERIENCE

Data Scientist Intern May 2023 - Aug 2023

Epsilon Data Management, LLC Chicago, IL, USA

- Analyzed the performance of the recommender system used by the company for different practical scenarios. The scenarios considered the time series behavior of the recommender, new user, item additions, and performance impact due to incremental training.
- Proposed AlBERT-based transformer architecture with multi-modal capabilities that considers user-personalized and user-item engagement features.
- Implemented CatBoost, a tree-based model which supports Apache Spark and Scala, and an AlBERT-based transformer model as potential solutions to improve the recommender used by the company.
- CatBoost and AlBERT-based transformer model results showed performance improvement over the recommender used by the company.
- CatBoost reduced the training time by 40% compared to the recommender used by the company.

Data Scientist Intern May 2022 - Aug 2022

Epsilon Data Management, LLC Chicago, IL, USA

- Developed a non-linear deep learning recommendation system that uses collaborative learning as a replacement for the current model. The proposed model generalizes well and can tackle continuous data without the need of external user, item or user-item features.
- The coding is done both in Scala and Python. BigDL library was used for Scala coding and coding in python was done using Tensorflow.
- The results demonstrated that the developed deep learning model was scalable and computationally faster in terms of train time and recommendation time for smaller companies.

Software Engineer Jun 2016 - Jan 2020

Oracle India Pvt. Ltd Hyderabad, TS, India

- Experience in Product Automation, Product Development and Product Deployment. Hand-on experience on production deployments and debugging.
- Knowledge of complete oracle stack like storage, servers, virtual machine, operating system, database, middleware and applications. Experience of automating complex product architectures.
- Worked on cloud-based platforms in hospitality industry related to hotel management.
- Extensively worked on Chef automation tool and coding was done in Ruby, Python, Bash and Java.

Intern Jun 2015 - Jul 2015

Rakshak Foundation New Delhi, India

- Worked on Unique identification number of India, aadhaar.
- Researched on the merits and demerits of using aadhaar in day-to-day activities.
- Proposed suggestions to improve effectiveness of aadhaar and steps to handle potential problems.

REFERENCES

Name	Title	Email	Department	University
Qi Li	Assistant Professor	qli@iastate.edu	Computer Science	Iowa State University
Oliver Eulenstein	Professor	oeulenst@iastate.edu	Computer Science	Iowa State University
Sihong Xie	Associate Professor	sihongxie@hkust-gz.edu.cn	Artificial Intelligence	HKUST(GZ)