Research Experience for Undergraduates (REU) Opportunity in **Data Analysis with Noisy Labels**

This opportunity to join Dr. Xiao Fu’s research group as an undergraduate student is possible through a research grant from National Science Foundation on “III: Small: Labeling Massive Data from Noisy, Incomplete and Crowdsourced Annotations” see <https://www.nsf.gov/awardsearch/showAward?AWD_ID=2007836&HistoricalAwards=false> for an abstract of the project.

**Scope of the project**: Alongside the prosperity of deep learning, the demand for reliably labeled data is unprecedentedly high. Label acquisition is a highly nontrivial task---data labeling is tedious, labor-intensive, and prone to mistakes. Crowdsourcing techniques that integrate annotations from multiple annotators to improve accuracy have been essential for labeling large-scale data. However, existing crowdsourcing techniques face pressing challenges such as heavy workload of annotators, high computational cost, and a lack of strong theoretical guarantees. This project will develop a series of analytical and computational tools for accurately labeling massive datasets from noisy, incomplete, and crowdsourced annotations---with provable guarantees. Leveraging advanced nonnegative matrix factorization theory, this project will offer solutions that are efficient and effective under critical conditions. The outcomes are expected to have broad and substantial positive impacts on the currently label-hungry artificial intelligence industry and the data annotation workforce. For example, the algorithms designed for handling structured data (e.g., speech) will largely benefit timely applications, e.g., intelligent assistants such as Alexa and Siri. The ability of reliably working under largely incomplete data will help design new data dispatch schemes leading to significantly reduced annotator workload. The project will also offer many training opportunities for undergraduate students, with an emphasis on engaging those from underrepresented groups.

**Qualification:** Students with junior and senior status will be given preference; however, highly-motivated sophomores with excellent academic achievements are also encouraged to apply. No prior background in the field is needed for you to apply.  A strong academic background and willingness to work hard and learn new things are the most important preparation for joining our team. Some mathematical tools that we will be using include: **1) Calculus 2) Linear algebra 3) Numerical Optimization 4) Signals and Systems 5) Probability and Random Variables.**

**What to be gained through the program**:

* Successful applicants will have the opportunity to work alongside with senior researchers for 10 weeks in the summer of 2022 to produce research outcomes under this NSF project.
* Valuable research experience benefiting future study and career is expected to be gained.
* An NSF-sponsored stipend will be available (up to $8000 for full time participation throughout the summer term).
* The research program can be extended to the next academic year depending on the group and the participant’s mutual interest.
* REU participants will be encouraged to co-author potential publications/research papers.

**Duties of the REU participant:**

* Performing daily research activities (paper reading, algorithm studying, programing)
* Participating research meetings
* Accomplishing research-related assignments, e.g., programming, algorithm coding/evaluation, data cleaning
* Writing progress reports and documenting research activities
* Presenting the research experience towards the end of the 10-week program

**How to Apply**: If you are interested in doing cutting edge research and developing technology that can substantially improve machine learning, please send an e-mail to Dr. Fu [xiao.fu@oregonstate.edu](mailto:mathews@oregonstate.edu) and Dr. Marrinan [marrinat@oregonstate.edu](mailto:marrinat@oregonstate.edu) to express your interest. Please include

* your resume
* a not-more-than-200-words statement on why you are an ideal candidate for this position with your application
* other materials that you think that could support your application, e.g., unofficial transcript and past research work

There is no hard deadline for the application, but candidates who send their materials before May 1st , 2022, will have higher chances of being selected. More information about Dr. Fu’s research group and their activities can be found from the websites of the group members:

<https://web.engr.oregonstate.edu/~fuxia/>

<http://www.tmarrinan.com/>

<https://shahanaibrahimosu.github.io/>

https://web.engr.oregonstate.edu/~lyuqi/