Yifan Yang

Department of Computer Science,
8125 Paint Branch Drive
College Park, MD 20742
Phone: 301-323-5740
Email: yang7832@umd.edu

Areas of Interest

Natural Language Processing, Algorithms, Machine Learning.

Education

Current — PhD in Computer Science,
University of Maryland College Park. Advisor: Vanessa Frias-Martinez

2020 — BS in Computer Science with honors,
University of Maryland College Park.

Research Experience

2020 - 2021 — BIKESHARE: E-bikes Effect on Mode and Route Choice: a Case Study of Richmond, VA , with Dr. Vanessa Frias-Martinez

This project aims to find the effect of electric-assist bikes (e-bikes) on the mode and route choice decisions of bikeshare users, providing recommendations on bike infrastructure, place and route guidance.

2018 - 2020 — Undergraduate Research Assistant in Dr. Héctor Corrada Bravo's lab,
Center for Bioinformatics and Computational Biology.

My research focuses on the development of computational packages for querying and analyzing genomic data directly from files. This involves parsing, caching and optimizing the index of genomic files for range queries. In addition, I worked on developing a scalable index to query genomic data from a large collection of files using Quadtree based approach.

2019 — Video OCR pipeline based on Mask R-CNN and shuffleNet (Google Summer of Code)

I worked on developing a machine learning pipeline to extract text from videos (mainly news). The workflow uses Mask R-CNN that takes shuffleNet as a backbone to detect text fragments, Tesseract OCR to detect text and dynamically merges these to eliminate duplicates and increase accuracy.

Publications

In submission

2021 Tahseen Rabbani, Brandon Feng, **Yifan Yang**, Arjun Rajkumar, Furong Huang. Comfetch: Federated Learning of Large Networks on Memory-Constrained Clients via Sketching

in preparation for submission

2021 **Yifan Yang**, Elliott Sloate, Nashid Khadem, Celeste Chavis, Vanessa Frias-Martinez. E-bikes' Effect on Mode and Route Choice: Clustering Causality and behavior patterns.

2021 (*Equal Contribution) Jayaram Kancherla*, **Yifan Yang***, Hector Corrada Bravo. Scalable index for accessing collections of functional genomic data using Quadtree.

Journals

2020 Jayaram Kancherla, **Yifan Yang**, Hyeyun Chae, Hector Corrada Bravo, Epiviz File Server: Query, Transform and Interactively Explore Data from Indexed Genomic Files, Bioinformatics, btaa591, https://doi.org/10.1093/bioinformatics/btaa591
Poster & Presentation at ISMB 2019.

Conference

Elliott Sloate, **Yifan Yang**, Nashid Khadem, Celeste Chavis, Vanessa Frias-Martinez. E-bikes' Effect on Mode and Route Choice: A Case Study of Richmond, VA Bike Share.

Research Projects

Current Detecting and Correcting Translation Artifacts in Multilingual NLI dataset

Both manual and machine translations add noise into the NLI dataset when creating NLI datasets of different language using the original dataset. These noises can corrupt the labels, accounting for bad performances of models. We aim to detect these corrupted examples in the dataset and correct them.

2021 Analyzing Multi-Head Self-Attention across different languages

Under the lottery ticket hypothesis, we test the performance of pruned models across different languages. Transferring to language that are less similar loses more accuracy. Positional attention heads and heads that put more attention to keywords are preserved to the last pruning iteration regardless of the language.

2020 Evaluating NLI dataset with generation approaches

Using Seq-to-Seq models, the method evaluates how much information that classifiers utilizes in prediction. The underlying assumption is that the more artifact in the NLI dataset, the less information classifiers are using.

2019 Quizbowl Question Generation Using Attention-based LSTM

we generate questions using open knowledge and quizbowl question set. The model encodes the knowledge input and translates the encoded meaning vectors to question (as sentences) with the help of an attention vector.

2019 R-CNN, Fast R-CNN, Faster R-CNN and Mask R-CNN: Techniques and Differences

A detailed report on how each neural network works and solves object recognition in images. I re-implemented and extended Mask R-CNN for the Google Summer of Code project.

2018 Segmenting Deformable Object from a Given Video Sequence

Tracing an object through continuous frames using probability color and shape model, this work is based on the Video Snapcut paper of Bai et al.

2018 Reconstructing 2-D Structures from Drone Captured Videos

Solves Simultaneous Localization and Mapping (SLAM) problem using GTSAM.