Ryan J. Cooper

Website : ryancooper.dev
Email : rjcooper@live.com

Ryan received his Masters of Science in Computer Science from Georgia Tech, focusing in Machine Learning and Natural Language Processing, and his Bachelors of Science in Computer Science from New Mexico Tech. He has 7+ years of practical experience in Machine Learning, Data Mining, and Software Engineering. He hopes to continue both research and applications in and around the field of Machine Learning.

Education

- **Georgia Institute of Technology (3.90/4)** — Atlanta, GA
  Masters of Science in Computer Science, Specialization in Machine Learning — *Aug. 2020 – Aug. 2021*
- **New Mexico Institute of Mining and Technology (4.00/4)** — Socorro, NM
  Bachelor of Science in Computer Science, Minor in Mathematics, Summa Cum Laude — *Aug. 2016 – May. 2020*
- **The ASK Academy** — Rio Rancho, NM
  Graduated as the Class of 2016 Valedictorian — *Aug. 2012 – May 2016*

Professional Experience

- **Sandia National Laboratories** — Albuquerque, NM
  Intern Year-Round R&D Computer Science, (Active DOE L Clearance) — *May 2017 - Present*
  - **Knowledge Engineering and Analytics**: Developed state of the art capabilities for customer projects leveraging Sandia's High Performance Computing environment and modern NLP techniques.
  - **Enterprise Search**: Worked on content discovery and organization of data from unstructured sources. Developed backup and recovery strategies for Enterprise Search.
  - **Search Analytics and Personalization**: Solved learning to rank problems using a deep neural network approaches using personalization.
- **Sandia National Laboratories** — Albuquerque, NM
  Intern Summer R&D Computer Science High School — *May 2015 - Aug. 2015*
  - **Research**: Worked as a Sandia STAR intern researching novel implementations of data mining in geospatial semantic spaces.

Technology Summary

- **Programming**: Python, C, Java, HTML, JavaScript, C#, C++, MATLAB, R, LaTeX
- **Machine Learning Technologies**: PyTorch, PyTorch Lightning, Tensorflow, Keras, NumPy, SciPy, OpenCV, Tesseract, spaCy, NLTK, Transformers
- **Software**: Kubernetes, Docker, Solr, Spark, ZooKeeper, Redis, Jenkins, MSSQL, Sqlite, MongoDB, Elasticsearch, Slurm, Github/Gitlab

Publications and Presentations

- **Configuring Recommendations for Personalized Search at Sandia National Laboratories (Activate Conference, 2019)**: In the scope of enterprise search, the assumed preference of each user is the number of times that they have previously clicked on pages, an observed weight. This weight is then used to co-cluster (associate) with other users to make predictions about what pages they will be most likely to find useful based on their previous click history.
- **Recurrent Neural Network Ranking System for Personalized Enterprise Search (OSTI, 2017)**: Machine-learned ranking (MLR) is a method of ranking search results based on relevancy to a query along with features about the search space. This project explores the process of implementing and evaluating a personalized MLR system in the scope of an Enterprise Search platform. By using search event information coupled with information about the user, a personalized approach to ranking pre-indexed results was developed and found to be effective.
- **Geospatial-Temporal Semantic Graph Search Template Generation via Data Mining (OSTI, 2015)**: Data mining plays a key role in search template generation for the analysis of large overhead image sets, particularly that of ontological storage, or geospatial-temporal semantic graph (GTSG). It provides an efficient method for determining the median of accuracy and consistency for template generation, one of which human analysts are required to provide substantial time and effort to create comparable results.

Awards and Achievements

1. Sandia National Laboratories Director Recognition - FY18 Top 3 Project Improvements (Jan 2019)
2. U.S. Presidential Scholars Program Semifinalist (May 2016)
3. NSA Stokes Program Finalist (April 2016)
4. New Mexico Science and Engineering Fair, First Place in Computer Science Division (March 2016)
5. Junior Science and Humanities Symposium National Competition Participant (March 2016)

Notable Projects

- **CP2R: GPT-2 Conversational Pipeline using Relevance and Realism Discrimination, (Georgia Institute of Technology, 2021)**: GPT-2 and it's successor GPT-3 have taken the open-text generation domain by storm having been leveraged in many open-domain conversational agents. We present CP2R, a GPT-2 based pipeline that attempts to overcome these pitfalls through a fine-tuning over 16 million posts and comments from the CasualConversation subreddit, where comments and posts are explicitly moderated for niceness. Our model generates candidate responses for a given text and picks the most appropriate and realistic one through a combination of outputs from our BERT-based realism discriminator and relevance discriminator
- **Towards a Machine Learning-based framework for Academic Paper Pre-evaluation, (Georgia Institute of Technology, 2021)**: The current review process of submissions to academic conferences is a manual task, typically done by fellow academics and requires long hours of reading. Initial pre-screening of papers must be done to determine if a paper's content is relevant to a particular conference's list of relevant topics. We demonstrate the feasibility of automatically detecting the topic of an academic paper based on its topic. Using unsupervised language models such as GloVE and BERT, we showed that there is significant linear separability of embeddings between in-conference samples and out-of-conference samples.
- **A Gait Speed Measurement System for the Clinical Workflow, (Georgia Institute of Technology, 2021)**: Gait speed has been proved to be a strong clinical indicator of mobility impairment in patients with neurological disorders and other related diseases. In this effort, we seek to define robust data pipelines and remove noise from real world patient data. We found that sensor and context fusion greatly reduced noise and we define two possible algorithmic approaches to accomplish the denoising and candidate matching problem. This work was apart of the Aware Home Research Initiative (AHRI) and will benefit from this research project in downstream tasks.
- **Conspiracy-BERT: A Pre-Trained Language Model for Conspiracy Theory Tweets, (Georgia Institute of Technology, 2020)**: We introduce Conspiracy-BERT, a transformer-based model pretrained on a corpus of over 12 million tweets related to various conspiracy theories such as QAnon, Flat Earth theory, and the COVID-19 "plandemic." Our model shows a relative improvement of over 23% when compared to its base model BERT-Large on the task of conspiracy stance detection in tweets. We believe our language model optimized on popular conspiracy content will be an invaluable tool for downstream NLP tasks related to conspiracies for content moderators and researchers looking to curb the spread of dangerous conspiracies online. Additionally, we contribute a large unlabeled data set of tweets related to 12 different conspiracies and a data set of 5000 conspiracy tweets labeled for stance detection.
- **A Semi-Supervised Approach for Image Labeling with Affinity Coding, (Georgia Institute of Technology, 2020)**: Modern machine learning techniques, especially in the domain of images owe much of their success to large quantities of labelled training data. However, they are not easy to obtain and hence not scalable. There is a need to automatically label datasets in an unsupervised manner by making use of only a small labelled set. GOGGLES[1] does that by proposing a novel way to obtain probabilistic labels for unlabelled image datasets automatically. They make use of pretrained networks to extract features at multiple scales and use similarities between images to cluster them. Then they assign labels using a small labelled set. We challenge various aspects of GOGGLES by replacing its core mechanisms with our approaches. We involve the development set to obtain the features for clustering by training an embedding layer. This led to a boost in accuracy by upto 18.6% from GOGGLES for the CUB dataset. We also added a mechanism to improve the accuracy with increase in the size of the labeled set, which was missing in GOGGLES.
- **Exploring the Capabilities and Possible Applications of Neural Turing Machine, (New Mexico Institute of Mining and Technology, 2019)**: The neural turing machines (NTM) is a class of learner that was introduced in 2014 by Google DeepMind. The NTM adds a "working memory" to the computational unit in a traditional artificial neuron, essentially causing the neuron to not only act on the input provided to the neuron, but also acting as a controller to its own working memory set. The NTM has been shown to not only solve turing-problems, but is hypothesized to be a super-turing approximation model [1]. The NTM has spawned significant research in memory-augmented computing and allows classical deep-learning to be applied to algorithmic processes. We explore open-source implementations of NTMs and analyze the extent of these capabilities, comparing these to the shortcomings of classical memory models like recurrent LSTM and GRU.