Piyush Chawla

Personal Website: https://pijusch.github.io/
Contact: +1 380-895-4401, Columbus Ohio
chawla.81@osu.edu
https://www.linkedin.com/in/chawla-piyush/
google-scholar

Interpretable AI & ML: Seeking internship for Summer-22

Education (The Ohio State University, Columbus OH)

Research: Interpretability, Visualization, Natural Language Proc, Deep Learning **GPA: 3.97/4.0**
Ph.D. Computer Science Engineering
Aug 2018 - Present

Technical Skills

Programming (Python, JavaScript, Java), **Libraries** (Pytorch, Tensorflow, Scikit Learn, Pandas, D3, Flask)
Key Courses Machine Learning, NLP, Optimization,High Performance Deep Learning, Visualization, Linear Algebra

Recent Projects

Log2Vec - AIOps IBM Research **Summer 2021**

- Developer-written logs guide users to understand and debug bugs in software services.
- Manually reading these logs is challenging given that these services could contain a myriad of micro services.
- Created a new logs dataset (~8 million unique logs) using public GitHub repositories.
- Trained log models like Doc2Vec, FastText and RoBERTa on these logs.
- Integrated these language models in IBM's proprietary anomaly detection pipeline.

Probing Static Word Embeddings for Relational Information **Spring 2021**

- Relation induction has been used to uncover relational information in static word embeddings.
- The existing works focus on linear models and only consider vector offset as the feature.
- We developed MLP based non-linear probes and found that features like vector product and sum contain significant amounts of information.
- Our results find differences between publicly available Skipgram and GloVe models.

Relation-Induction on Neural Word Embeddings *(Python, Pytorch)* **Fall 2020**

- Word embeddings contain a myriad of information and possibly more that still needs to be uncovered.
- Developed classification models for learning semantic relationships between words.
- Beat SOTA on relation datasets BATS, Google, DiffVec using Skip-Gram (Word2vec) and Glove embeddings.
- Created few-shot learning models for limited-data variants of the classification models.
- Implemented meta-learning algorithms MAML (FOMAML) and Reptile.

Understanding Convolutional Neural Networks for text *(D3, Javascript, Python, Flask)* **Spring 2020**

- ConvNet visualization has been explored for computer vision. However, understanding this class of architecture still remains an open problem in the context of natural language (text) applications.
- Developed an approach (TSD) to visualize the contribution (+ve/-ve) of each word in a sentence towards the predicted label (sentiment).
- Discussed possible adversarial and error analysis strategies for ConvNet-based sentiment-analysis classifiers.

Publications/Ongoing Works

1. **P. Chawla**, Y. Su, H.W. Shen (2021) Relation induction on word embeddings: A few-shot learning approach. (under review)
2. H. Choi, **P. Chawla**, H.W. Shen (2021) Topic Tracking for Time-Varying Text Data. (name changed for blind-review)
3. D. Esteves, J. Marcelin, **Piyush Chawla**, A. Fischer, J. Lehmann (2021) HORUS-NER: A Multimodal Named Entity Recognition Framework for Noisy Data. IDA 2021
4. **Piyush Chawla**, S. Hazarika, HW Shen (2020) Token-wise sentiment decomposition for ConvNet: Visualizing a sentiment classifier. PacificVis 2020
5. **Piyush Chawla**, D. Esteves, K. Pujar, J. Lehmann (2019) SimpleLSTM: A Deep-Learning Approach to Simple-Claims Classification. EPIA-2019
6. D. Esteves, A. J. Reddy, **Piyush Chawla** and J. Lehmann (2018) Belittling the Source: Trustworthiness Indicators to Obfuscate Fake News on the Web. EMNLP 2018

Research/Work Experience

PhD, The Ohio State University, Columbus, OH **Aug 2018 - Present**

- GRAVITY Lab (Dr. Han-Wei Shen)
- DKI Lab (Dr. Yu Su)
- Graduate Teaching Assistant

Research Intern, IBM Almaden, San Jose (Virtual) **May 2021 - Aug 2021**

Project: Log Anomaly Detection

- Generated log dataset from millions of publicly available GitHub repositories.
- Designed downstream tasks to gauge the quality of the log dataset.
- Trained language models like Doc2Vec and RoBERTa.
- Plugged-in these language models into IBM's proprietary log anomaly detection pipeline.

Research Assistant, Smart Data Analytics, Bonn-Germany **Dec 2017 - June 2018**

Bachelor's Thesis: Link prediction in Multi-Lingual Knowledge Graphs (Python, Tensorflow)

- Used interlanguage links (e.g. English-German) in DBpedia knowledge graph.
- Implemented the mTransE model to improve link prediction accuracy.
- Developed a model (NLTransE) to add semantic knowledge from word embeddings to KG embeddings.

Summer Intern, University of Bonn DAAD WISE Scholarship (Top-100 all-over India) **May 2017 - Jul 2017**

Project: Knowledge graph completion using latent vector models (Python, Tensorflow, C++)

- Explored different latent-vector models for KG link prediction. TransE, TransR, TransH, DistMult etc.
- Conducted large-scale experiments on DBpedia Knowledge Base to train KG completion models.

Professional Service

Secondary Reviewer

IEEE TKDE 2019, KDD 2020, IEEE VAST, ACL 2021, ISVC 2021

Awards and Achievements

- University Fellowship, The Ohio State University - 2018-2019
- Bachelor's Thesis Support, BITS Pilani - 2017-2018
- Thesis Research Support, University of Bonn - 2017-2018
- WISE scholarship, Deutscher Akademischer Austauschdienst - Summer 2017
- Merit scholarship, BITS Pilani - 2014-2018