

# Piyush Chawla

Deep Learning, Machine Learning

Personal Website: <https://pijus.ch.github.io/>

Contact: +1 380-895-4401, Columbus Ohio

✉ [chawla.81@osu.edu](mailto:chawla.81@osu.edu)

in <https://www.linkedin.com/in/chawla-piyush/>

g [google-scholar](#)

## Education

**Research: Applied Machine Learning, Model Interpretability, Natural Language Processing**

**The Ohio State University**

Master of Science, Computer Science Engineering (**Transferred from Ph.D. post candidacy**)

GPA: 3.98/4.0  
2018 - Dec 2022

**Birla Institute of Technology & Science**

Bachelor of Engineering (Hons), Computer Science Engineering

GPA: 9.7/10.0  
2014 - 2018

## Technical Skills

**Programming** Python, Pytorch, Hugging-Face, Torch-Geometric, Git, Version Control, Scikit Learn, SciPy, Gensim

**Key Courses** Machine Learning, NLP, AI, Data Mining, Optimization, High Performance (Distributed) Deep Learning

**Misc.** BERT, LSTM, CNN, RoBERTa, Transformer, GNN, GCN, GGNN, GRU, Sequence Modeling, Mixture Models

## Research/Work Experience

**PhD Student, The Ohio State University, Columbus, OH**

Aug18 - Now

- GRAVITY Lab (Dr. Prof. Han-Wei Shen)
- Graduate Assistant

**Research Scientist Intern, Bosch Research, Sunnyvale, California**

May - Aug 22

*Project: Explainable Graph Similarity for Workflow Graphs*

- Developed XAI solution for the problem of Graph Similarity on PascalVOC image key-point graphs.
- Leveraged Self Supervised Learning (SSL) to achieve new SOTA on node alignment based explainability.
- Result: Proposed novel structure-centric SSL pretext tasks and multi-task learning (MTL) based ML framework.
- Result: Filed research patent for the idea.

**Research Intern, IBM Almaden, San Jose, California**

May - Aug 21

*Project: Log Anomaly Detection*

- Generated a large scale (8M logs) high quality dataset by parsing publicly available GitHub repositories.
- Designed downstream tasks to gauge the quality of the log dataset.
- Trained language models Doc2Vec and RoBERTa for log anomaly detection tasks achieving new SOTA.
- Result: Plugged-in these language models into IBM's proprietary log anomaly detection pipeline.

**Research Assistant, Smart Data Analytics, Bonn-Germany**

Dec 17 – July 18

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

- Leveraged interlanguage links (e.g., English-German) in DBpedia knowledge graph to train KG embeddings.
- Implemented the mTransE model to improve link prediction accuracy.
- Result: Novel method dubbed *NLTransE* to transfer word embedding semantic knowledge to KG embeddings.

**Intern (Research Scholar), University of Bonn DAAD WISE (1 of the 100 selected all over India)**

May – Jul 17

*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.
- Result: Conducted large-scale experiments on DBpedia Knowledge Base to train KG completion models.

## Key Projects

**Question Answering On Table Data (Python, Pytorch)**

Fall 2022

- Working to develop novel Question Answering method for hybrid data (structured and unstructured).

**Pattern Recognition in Academic Publications (Python, Pytorch, D3, JavaScript)**

Spring 2022

- Clustering of text document is commonly used in pattern recognition but explaining the clusters is a challenge.
- Proposed a novel clustering approach in topic model latent space using Mixture of Gaussians.
- Used resulting clusters to find interesting “research themes” emerging in the academic publications.
- Result: A novel visualization system to discover temporal trends in documents. (publication under review)

### Probing Static Word Embeddings for Relational Information (*Python, Pytorch*)

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. Results: Achieved new SOTA on relation induction problem.
- Findings: Publicly available GloVe embeddings contain richer information than Skipgram.
- Results: Discussed extensions for bias detection static word embeddings. (publication under review)
- As an extension, we train models in meta learning setting (MAML, Reptile), improving SOTA few-shot prediction.

### Understanding Convolutional Neural Networks for text (*D3, JavaScript, Python, Flask*)

Spring 2020

- CNN visualization (explanation) has been explored for computer vision. However, understanding this class of architecture remains an open problem in the context of natural language (text) applications.
- Result: Developed a novel approach dubbed “*Token-wise Sentiment Decomposition*” to visualize the contribution (+ve/-ve) of each word in a sentence towards the predicted label (sentiment).
- Result: Discussed adversarial and error analysis strategies for ConvNet-based sentiment-analysis classifiers.

### Finished Publications

- Piyush Chawla, S. Hazarika, HW Shen (2020) Token-wise sentiment decomposition for ConvNet: Visualizing a sentiment classifier. PacificVis 2020
- Piyush Chawla, D. Esteves, K. Pujar, J. Lehmann (2019) SimpleLSTM: A Deep-Learning Approach to Simple-Claims Classification. EPIA-2019
- D. Esteves, J. Marcelin, Piyush Chawla, A. Fischer, J. Lehmann (2021) HORUS-NER: A Multimodal Named Entity Recognition Framework for Noisy Data. IDA 2021
- 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

### Awards and Achievements

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

### Professional Service

*Secondary Reviewer*

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