Neeraj Varshney

Ph.D. Student (Third Year)
Computer Science (NLP)
Arizona State University, CGPA: 4.1/4

Email: nvarshn2@asu.edu
LinkedIn: neerajvarshney97
Mob: +1(480)561-9416
Tempe, Arizona

Profiles

LinkedIn: **neerajvarshney97**
Scholar: **Ju9nR0IAAAAJ&hl**
Medium: **@nvarshney97**
Github: **nrjvarshney**

Education

Ph.D. Computer Science (NLP)
Arizona State University, USA
2019-2024 EXPECTED
CGPA: 4.1/4

B.E.(Hons.) Computer Science
Birla Institute of Technology and Science (BITS Pilani, India)
2014-2018
CGPA: 9.11/10

Coursework

Natural Language Processing
Statistical Machine Learning
Artificial Intelligence
NLP Methods in BioMedical
Knowledge Representation & Reasoning, Data Mining

Technical Skills

ML, DL, PyTorch, Transformers, Pytorch-lightning, Python, Jupyter, Pandas, Git, GitHub, Google Colab, Spacy, Linux, NumPy, Matplotlib, Huggingface, NLTK, word2vec

OTHERS

- Worked with Dr. Ayush Choure (MSR) in a project lead by Dr. Prateek Jain (MSR).
- Published 10+ ML/NLP related articles on medium with 2000+ monthly views.
- Organized 6^{th} edition of Alumni Research Talks being the Campus Coordinator of Computer Science Association at BITS.
- Worked at "Web Intelligence & Social Computing" lab under Prof. Poonam Goyal at BITS.
- Won 2nd prize in Technical fest prototype project at BITS Pilani.

Publications

ILDAE: Instance-Level Difficulty Analysis of Evaluation Data *ARR, SEP, 2021*
Conducted Instance-Level difficulty analysis in a large-scale setup of 23 datasets with 27 models and demonstrated its five novel applications such as:
- Efficient Evaluations: Proposed an instance selection technique that achieves 0.72 Kendall correlation with full dataset evaluation using just 20% instances.
- Dataset Quality: Proposed a model-and-human-in-the-loop technique that modifies/repairs trivial and erroneous instances to improve the quality of evaluation datasets.

Unsupervised NLI Using PHL Triplet Generation *PREPRINT, 2021*
- Designed three novel unsupervised settings for NLI and proposed a procedural data generation approach that outperforms the existing approaches by $\sim$13% and raises the SOTA unsupervised performance to 66.75% on the SNLI dataset.
- Provided a strategy to efficiently collect high-quality task-specific data that helps achieve 12.2%, 10.4% higher accuracy on SNLI and MNLI with just 500 human-authored instances.

It's Better to Say "I Can't Answer" than Answering Incorrectly *PREPRINT, 2020*
- Demonstrated that MaxProb as a selective answering technique performs well on In-Domain inputs but fairs poorly on Out-of-Domain inputs.
- Proposed a novel selective answering approach that incorporates prediction confidence and instance-level difficulty score to calibrate model's outputs and achieves an improvement of up to 7.47% over existing methods on AUC of risk-coverage curve.

Let the Model Decide its Curriculum for Multitask Learning *PREPRINT, 2020*
- Proposed two classes of techniques (dataset and instance-level) to arrange training instances into a learning curriculum based on model's own interpretation of difficulty.
- Achieved 4% accuracy improvement over other methods on experiments conducted for 12 datasets covering varied sentence pair tasks such as NLI, duplicate detection.

NumGLUE: A Suite of Mathematical Reasoning Tasks *ARR, AUG, 2021*
- Proposed a multi-task benchmark that evaluates the performance of AI systems on eight different tasks that require simple numerical understanding.
- Showed that this benchmark is far from being solved with neural models including large language models performing significantly worse than humans (lower by 46.4%).

Interviewer-Candidate Role Play: Towards Real-World NLP Systems *PREPRINT, 2021*
- Designed a multi-stage selective answering task that assists a model with instance-related information such as knowledge statements, examples at various stages when the model is not sufficiently confident in its prediction (post-abstention scenarios).
- Achieved improvements of up to 72.02% in Out-of-Domain generalization.

Can Transformers Reason About Effects of Actions? *PREPRINT, 2020*

Ongoing Projects

- SelectiveBench: A suite of tasks to evaluate selective prediction ability of systems on in-domain, out-of-domain, adversarial, and unanswerable inputs in four different settings.
- Compare selective prediction ability of Autoregressive, Autoencoding, and Sequence-to-sequence models that achieve similar accuracy.
- Learning from instructions: Investigating the impact of conditioning large language models like GPT-3 on task instructions and a few examples in natural language format.

Experience

Microsoft *JAN 2018 - JULY 2019* **Software Engineer**
- Contributed towards development of a Machine Learning driven chat recommendation system aimed at augmenting user engagement with Microsoft's product 'Teams'.
- Collaborated with MSR researchers for a feature titled 'Intelligent Feeds' that finds relevant messages for users based on their prior activities and message text features.