Barun Patra

E-Mail : barunpatra95@gmail.com | Github : codededcde@github.com | Mobile No. : (425) 465 7261
Website: https://www.microsoft.com/en-us/research/people/bapatra
Google Scholar: https://scholar.google.com/citations?hl=en&user=Gwg25AkAAAAJ

Research Interests

My research interest focuses on pre-training large models for multi-modal and multilingual applications, with an emphasis on inference speed, novel abilities like large context understanding and alignment to human preference. My work revolves around all aspects of pre-training: scaling data and models, exploring novel model variants for improving inference speed, sparse scaling and enabling larger context processing, as well as post-training alignment research. I am also interested in the limitations of such large models, and improving their robustness and interpretability.

Education

Carnegie Mellon University, Pittsburgh August 2017 - December 2018
Master's of Science in Machine Learning
Cumulative Grade Point Average: 4.17

Indian Institute of Technology, Delhi July 2013 - June 2017
B.Tech in CS&E with a Specialization in Data Analytics and AI
Cumulative Grade Point Average: 9.494/10

Work Experience

Turing Team: Microsoft Vancouver, Microsoft Redmond Feb 2022 - Present, Nov 2020 - Feb 2022
Link to Team Description *Senior Applied Scientist, Applied Scientist II*
I work on building better representation and generative foundational multilingual and multimodal language models. The work is instrumental in pushing frontiers on academic benchmarks (1^{st} on both XTREME and GLUE benchmarks), as well as cutting edge user experiences like Bing Chat. My work spans all aspects of building these models: processing petabyte scale data, scaling models to billions of parameters, exploring novel ways to reduce latency, developing different pre-training strategies as well as post-training strategies to improve the alignment of these models. I also work closely with different university research labs further advance research on understanding the capabilities and shortcomings of these models (Microsoft Turing Academic Program).

Scheduler Team, Microsoft Redmond May 2018-Aug 2018, Feb 2019 - Nov 2020
Link to Team Description *Applied Scientist Intern, Applied Scientist*
I developed NLP components for email scheduling assistant, including intent classification, entity extraction, time understanding, negation handling and understanding spans of interest in emails. This consequently improved automation levels from below 20% to over 90% and contributed to 2 publications and 4 patents. I also organized NLP workshops and reading groups.

Masters in ML, CLU Aug 2017 - Dec 2018
During my Masters, I had the opportunity to work with Professor Matthew Gormley, Professor Barnabàs Póczos and Professor Graham Neubig. Noteworthy projects include Knowledge Based completion using an RL agent with Monte Carlo Tree Search for generating supervised policies for the agent, where the resulting method yielded state-of-the-art results on 3 popular KB datasets, model compression for a DQN network (down to 3% of the original parameters while retaining performance) and quantifying degree of mis-alignement between embedding spaces of different languages and providing a semi-supervised framework for aligning embedding spaces of different languages.

DAIR, IIT Delhi July 2016 - July 2017
I was a part of the Data Analytics and Intelligence Research Group lab and was advised by Professor Mausam and Professor Parag Single. During my time there, I worked on incorporating global logical constraints to CRF inference for tagging entities of interest and using it for semi-supervised learning for answering multi-sentence tourism related questions. I also worked on automated dialogue system evaluation, leveraging the thread structure of Reddit to generate context, response and alternate response triples, with the final model achieving a zero-shot Pearson correlation of 0.45 with human scores on the Ubutnu dataset.

Teaching Experience

Advanced Introduction to ML (CMU) Fall 2018
Teaching Assistant under the guidance of Professor Nina Balcan
Worked on developing questions for assigments and exams, graded exams, provided mentorship and recitations for 70+ primarily PhD students

Artificial Intelligence (IIT Delhi) Spring 2017
Teaching Assistant under the guidance of Professor Mausam
Graded exams and quizzes and developed assigments for 125+ students in this grad/undergrad bridge course

Academic Services

I have served as a reviwer for the following conferences

ACL Rolling Reviews (2021, 2022, 2023)	NeurIPS (2022, 2023)	GHC (2020)
EMNLP (2022†, 2023)	ACL(2021, 2022)	
TPAMI (2023)	ICLR (2022)	AAAI (2018)

† *indicates outstanding reviewer*

Patents

Patra Barun, Fufa Chala, Bhattacharya Pamela, Lee Charles. Identifying contextual and task relevant time entities with constraints (MS# 408676-US-NP), filed on 09/18/2020

Suryanarayanan Vishwas, Patra Barun, Bhattacharya Pamela, Fufa Chala, Lee Charles. Artificial Intelligence For Identifying Relevant Content Related To Specific Tasks (MS# 407907-US-NP), filed on 12/06/2019

Patra Barun, Bhattacharya Pamela, Lee Charles. Resolving Temporal Ambiguities in Natural Language Inputs Leveraging Syntax Tree Permutations (MS# 407193-US-NP), filed on 9/16/2019

Patra Barun, Bhattacharya Pamela, Lee Charles. Sentence Attention Modeling For Event Scheduling via Artificial Intelligence and Digital Assistants (MS# 405393-US-NP), filed on 11/30/2018

Publications and Preprints

Everything you need to know about Multilingual LLMs: Towards fair, performant and reliable models for languages of the world Sunayana Sitaram, Monojit Choudhury, Barun Patra, Vishrav Chaudhary, Kabir Ahuja, Kalika Bali. (**Tutorial @ ACL 2023**)

Language Is Not All You Need: Aligning Perception with Language Models Shaohan Huang, Li Dong, Wenhui Wang, Yaru Hao, Saksham Singhal, Shuming Ma, Tengchao Lv, Lei Cui, Owais Khan Mohammed, Barun Patra, Qiang Liu, Kriti Aggarwal, Zewen Chi, Johan Bjorck, Vishrav Chaudhary, Subhojit Som, Xia Song, Furu Wei. (**arXiv:2302.14045**)

Beyond English-Centric Bitexts for Better Multilingual Language Representation Learning Barun Patra*, Saksham Singhal*, Shaohan Huang*, Zewen Chi, Li Dong, Furu Wei, Vishrav Chaudhary and Xia Song. (**ACL 2023**)

A Length-Extrapolatable Transformer Yutao Sun, Li Dong, Barun Patra, Shuming Ma, Shaohan Huang, Alon Benhaim, Vishrav Chaudhary, Xia Song, Furu Wei. (**ACL 2023**)

Foundational Transformers Hongyu Wang, Shuming Ma, Shaohan Huang, Li Dong, Wenhui Wang, Zhiliang Peng, Yu Wu, Payal Bajaj, Saksham Singhal, Alon Benhaim, Barun Patra, Zhun Liu, Vishrav Chaudhary, Xia Song and Furu Wei. (**ICML 2023**)

Language Model Decoding as Likelihood-Utility Alignment Martin Josifoski, Maxime Peyrard, Frano Rajic, Jiheng Wei, Debjit Paul, Valentin Hartmann, Barun Patra, Vishrav Chaudhary, Emre Kiciman, Boi Faltings, Robert West. (**EACL 2023**)

TorchScale: Transformers at Scale Shuming Ma, Hongyu Wang, Shaohan Huang, Wenhui Wang, Zewen Chi, Li Dong, Alon Benhaim, Barun Patra, Vishrav Chaudhary, Xia Song, Furu Wei. (**arXiv:2211.13184**)

The SUMEval 2022 Shared Task on Performance Prediction of Multilingual Pre-trained Language Models Kabir Ahuja, Antonios Anastasopoulos, Barun Patra, Graham Neubig, Monojit Choudhury, Sandipan Dandapat, Sunayana Sitaram, Vishrav Chaudhary. (**AACL 2022**)

On the Representation Collapse of Sparse Mixture of Experts Zewen Chi, Li Dong, Shaohan Huang, Damai Dai, Shuming Ma, Barun Patra, Saksham Singhal, Payal Bajaj, Xia Song, Xian-Ling Mao, Heyan Huang, Furu Wei. (**NeurIPS 2022**)

Invariant Language Modeling Maxime Peyrard, Sarvjeet Singh Ghotra, Martin Josifoski, Vidhan Agarwal, Barun Patra, Dean Carignan, Emre Kiciman, Robert West, (**EMNLP 2022**)

On Efficiently Acquiring Annotations for Multilingual Models Barun Patra*, Joel Ruben Anthony Moniz*, Matthew R Gormley. (**ACL 2022**)

ScopeIt: Scoping Task Relevant Sentences in Documents: Barun Patra*, Vishwas Suryanarayanan*, Pamela Bhattacharya, Chala Fufa, Charles Lee. Proceedings of the 28th International Conference on Computational Linguistics (**COLING 2020, Industry Track**)

To Schedule or not to Schedule: Extracting Task Specific Temporal Entities and Associated Negation Constraints: Barun Patra, Chala Fufa, Pamela Bhattacharya and Charles Lee. (**EMNLP 2020**)

Constrained BERT BiLSTM CRF for Understanding Multi-Sentence Entity Seeking Questions: Danish Contractor, Barun Patra, Mausam and Parag Singla. (**JNLE 2020**)

Weakly Supervised Attention Networks for Entity Recognition: Barun Patra* and Joel Ruben Anthony Moniz*. (**EMNLP-IJCNLP 2019**)

Bilingual Lexicon Induction with Semi-Supervision in Non-Isometric Embedding Spaces: Barun Patra*, Joel Ruben Anthony Moniz*, Sarthak Garg*, Matthew R Gormley and Graham Neubig. (**ACL 2019**)

Understanding complex multi-sentence entity seeking questions: Danish Contractor, Barun Patra, Mausam and Parag Singla. (**AAAI Reasoning for Complex Question Answering Workshop, 2018**).

Compression and Localization in Reinforcement Learning for ATARI Games: Joel Ruben Anthony Moniz*, Sarthak Garg* and Barun Patra*. (**NeurIPS Deep Reinforcement Learning Workshop, 2018**)

A Survey of Community Question Answering Barun Patra. (**arXiv:1705.04009**)

* denotes equal contribution

Scholastic Achievements

Department Rank 3 : Top 3% among students of Computer Science and Engieering Department, Batch of 2017

Suresh Chandra Memorial Trust Bachelor's Thesis Award (2017) : Best B.Tech Project

Narotam Sekhsaria Postgraduate Scholarship (NSF Foundation) (2017) : Selected as one of the top 18 students in India pursuing Pure and Applied Sciences, Social Sciences, Law, Architecture and Management.

KC Mahindra Scholarship (2017) : Selected as one of the top 50 students in India for pursuing post-graduate studies abroad in varied fields

Merit Scholarship For being in the top 7% for 5 semesters.