Abhishek Panigrahi

Research Fellow, Microsoft Research India
Webpage ◇ Google Scholar ◇ Email

EDUCATION

- **Indian Institute of Technology, Kharagpur** July'14 – May'18
  B.Tech in Computer Science and Engineering
  Cum. GPA: 9.90/10, Major GPA: 10/10, Institute Rank : **1 (Out of 1400 students)**

WORK EXPERIENCE

- **Microsoft Research India** July'18 – Present
  Research Fellow
  · *Advisors: Dr. Harsha Vardhan Simhadri* and *Dr. Navin Goyal*.
  · Projects - Unsupervised Embeddings and Analysis of deep learning algorithms.
- **University of Southern California** May'17 – July'17
  Research Intern
  · *Advisor: Dr. C.-C. Jay Kuo*.
  · Project - Mathematical model for gradient back propagation in batch normalized models.

PUBLICATIONS

- Effect of Activation Functions on the Training of Overparametrized Neural Nets *[arxiv]*
  Abhishek Panigrahi, Abhishek Shetty and Navin Goyal
  (Under review) International Conference on Learning Representations. (ICLR 2020)
- Non-Gaussianity of Stochastic Gradient Noise *[arxiv]*
  Abhishek Panigrahi, Raghav Somani, Navin Goyal and Praneeth Netrapalli
  Science meets Engineering of Deep Learning workshop, NeurIPS 2019 (SEDL 2019)
- Word2Sense: Sparse Interpretable Word Embeddings *[paper]*
  Abhishek Panigrahi, Harsha Vardhan Simhadri and Chiranjib Bhattacharyya
  Association for Computational Linguistics (ACL 2019)
- Analysis on gradient propagation in batch normalized residual networks *[arxiv]*
  Abhishek Panigrahi, Yueru Chen, C.-C. Jay Kuo
- DeepTagRec: A Content-cum-User Based Tag Recommendation Framework for Stack Overflow *[paper]*
  Suman Kalyan Maity, *Abhishek Panigrahi*, Sayan Ghosh, Arundhati Banerjee, Pawan Goyal, Animesh Mukherjee
  European Conference on IR Research (ECIR 2019)
- Book Reading Behavior on Goodreads Can Predict the Amazon Best Sellers *[paper]*
  Suman Kalyan Maity, *Abhishek Panigrahi* and Animesh Mukherjee
  ACM International Conference on Social Networks Analysis and Mining (ASONAM 2017)

RESEARCH EXPERIENCE

- **Non-Gaussianity of Stochastic Gradient Noise** July'19 – Present
  Advisors: Dr. Navin Goyal *and* Dr. Praneeth Netrapalli*, Microsoft Research India*
  · Experimentally studied the distribution of the Stochastic Gradient Noise (SGN) vectors, at different batch sized stochastic gradient descent and different hyperparameter settings for different data-sets and architectures. Observed that for batch sizes 256 and above, the distribution is best described as Gaussian at-least in the early phases of training. However, for batch sizes 256 and below, the distribution is not Gaussian either in later phases of training or the entire training.

- Analysing the relation between the difference in distribution of SGN vectors and generalization gap between small batch and large batch training.
- Initial manuscript accepted in *SEDL workshop*, *NeurIPS 2019*.

- **Effect of Activation Functions on the Training of Overparametrized Neural Nets** July'18 – May'19
  Advisor: Dr. Navin Goyal, *Microsoft Research India*
  - Analyzed activation functions' impact on training time of 2-layer overparameterized neural networks.
  - Provided worst case polynomial convergence bounds, in number of samples, for functions with jump discontinuity e.g. ReLU,ELU, etc. Also, showed a regime, dependent on the dimension of data, where convergence bounds have exponential dependence on number of samples for smooth functions e.g. polynomials, swish, tanh etc. Also, showed that the bounds on the data dimension were tight. Showed that no such bad regime exists for sufficiently deep networks.
  - This work is under submission at *ICLR 2020*.

- **Word2Sense: Sparse Interpretable Word Embeddings** July'18 - March'19
  Advisors: Dr. Harsha Vardhan Simhadri *and* Dr. Chiranjib Bhattacharyya, *Microsoft Research India*
  - Proposed an unsupervised method to generate sparse and interpretable single prototype word embeddings, that contains explicit information about all the senses of a word.
  - Proposed an optimization method, that uses sparsity of Word2Sense, to explicitly specify the meaning of a word in a context. Performs the best in word sense disambiguation task.
  - The work got accepted for oral presentation at *ACL 2019*.

- **Regularization of GANs** Sept'17 – May'18
  Bachelor Thesis, Advisor: Dr. Pabitra Mitra, *IIT Kharagpur*
  - Worked on regularizing GANs, that use wasserstein loss or maximum mean discrepancy loss, to solve multi-manifold problem.
  - One approach was to learn tangent space at each point by local PCA and match them, since points in generated manifold and original manifold that are closer should have similar tangent planes.
  - Another approach was based on boosting to increase the weights of generated points not present in original manifold and construct a weighted MMD formulation using those weights. In low dimensional data with multiple independent clusters, MMD (wasserstein) GANs give interconnected clusters as output, while weighted MMD was (partially) successful in separating them.

- **Mathematical model for gradient back propagation in batch normalized models** May'17 – Oct'17
  Advisor: Dr. C.-C. Jay Kuo , *University of Southern California*
  - Developed a complete mathematical model for modeling the gradient back propagation in batch normalized models and residual networks.
  - Showed that the residual branch and the batch normalization in each residual block, stop the gradient from vanishing as well as exploding, thus maintaining variance of gradients across a batch flowing through the model, atleast in the initial phases of the training. Supported theory by numerous experiments, conducted on residual networks. Hence, residual branches and batch normalization stabilize learning.

- **Reading Behavior Analysis on Goodreads.com** May'16 – Dec'16
  Supervised by Dr. Animesh Mukherjee, *IIT Kharagpur*
  - Predicted whether a book will become a best seller. Extracted features of a book from Goodreads.com by using sentiment analysis from reviews, calculating uniqueness in book description and genre and shelf diversity (using KL divergence, cross entropy) and analyzing author details (in a similar way).
  - We used SVM and logistic regression classifier on the features extracted. The proposed prediction framework achieved a very high avg. accuracy of 88.72% with high avg. precision and recall (0.887) and achieved an improvement of 16.4% over the traditional popularity factors (ratings, reviews) based baseline methods. This work was accepted in *ASONAM 2017*.

ACADEMIC ACHIEVEMENTS

- **President of India Gold Medal 2018** Awarded for the best GPA in the 2014 batch
- **Department Silver medal 2018** Awarded for the best GPA in the 2014 computer science department batch
- **Viterbi India program 2017** Awarded to 20 students from India for funding their summer internship at University of Southern California, Los Angeles

- **Goralal Singhal Scholarship 2016-2017, J.C. Ghosh Memorial Endowment prize 2017, John Von Neuman Award 2017, R.M. Lalwani Award 2017, C. Devi Memorial prize 2017** For academic excellence
- **IIT-JEE Advanced 2014 : AIR 277** Among 1,50,000 students from across the country.
- **JEE Main 2014 : AIR 297** Among 1.5 Million students from across the country.
- **KVPY Fellow 2012** Scholarship by the Department of Science and Technology, Government of India.

RELEVANT COURSES

• Advanced Machine Learning • Advanced Linear Algebra • Foundations of cryptography • Operations Research • Probability and Statistics • Matrix Algebra • Deep Learning • Machine Learning + Practicum • Natural Language Processing + Practicum • Artificial Intelligence • Computer Graphics + Practicum • Image Processing* + Practicum • Discrete Structures • Algorithms I + Practicum • Algorithms II

REFERENCES

- *Dr. Navin Goyal*, Microsoft Research India
- *Dr. Harsha Vardhan Simhadri*, Microsoft Research India
- *Dr. C.-C. Jay Kuo*, Electrical Engineering, University of Southern California
- *Dr. Pabitra Mitra*, Computer Science and Engineering, IIT Kharagpur