Anurendra Kumar

217-819-8065 • Champaign, Illinois 61820 • ak32@illinois.edu • anurendra.github.io

RESEARCH INTERESTS: Computational Biology, Machine Learning

EDUCATION

University of Illinois at Urbana-Champaign Aug '19 - May '24
M.S., Ph.D. in Computer Science (Concentration in Bioinformatics) GPA: 4/4

Indian Institute of Technology (IIT), Kanpur Jul '12 - May '17
M. Tech. in Electrical Engineering GPA: 9.67/10
B. Tech. in Electrical Engineering (Minors in Artificial Intelligence, Linguistics) GPA: 8.0/10

EXPERIENCE

Graduate RA, Prof. Saurabh Sinha, UIUC Dec '19 - Present
Single cell genomics
• *Spatio-temporal Transcritomics*: Proposed statistical approaches to understand the spatio-temporal regulation of genes. Developed a hypothesis testing with a realistic null model, which outputs a subset of spatially co-localized genes
• *Gene Regulatory Network (GRN)*: Developed a parameter estimation for single cell simulator with an underlying GRN. Proposed a transfer learning framework, which can utilize synthetic along with real data for optimal learning
• *Neurogenomics*: Proposed an attention based deep neural network (DNN) to i) decode combinatorial gene regulation and ii) learn relevant motifs from the enhancer sequences

Machine Learning Engineer Intern, Service Now May '20 - Aug '20
• Developed a DNN using **BERT** for content extraction from document images which yielded ∼**85**% accuracy on SROIE dataset

Graduate RA, Prof. Kevin Chang, UIUC Aug '19 - May '20
Context-aware Webpage Object Detection [Submitted to CVPR '21] **[Code]**
• Proposed an attention based DNN which yielded interpretable results and achieved ∼**10%** improvement over SOTA
• Created **largest public labeled dataset of** $7.7k$ **product webpage screenshots**

Lead, Data & Artificial Intelligence, Startup Project Jan '18 - Mar '19
• Developed algorithms and benchmark metrics for various finance services e.g. credit scoring & risk estimation

Research Associate, IIT Kanpur-ISRO Collaboration Aug '17 - Dec '17

Research Intern, IBM Research Lab Jun '17 - Aug '17
Hierarchical sparse representation of Knowledge base (KB) [**Presentation**]
• Proposed a tree structured prior for representation learning of KB. Used proximal gradient to deal with non-smoothness

Extreme Blue Intern, IBM May '16 - Jul '16
Internet of Things (IOT) in Agriculture **[Video]**
• Crop health monitoring from NIR & RGB images and various sensors. Awarded **AEGIS GRAHAM BELL AWARD**

Research Intern, Samsung May '15 - Jul '15

PUBLICATIONS

- **A.Kumar**, T.Guha, P.Ghosh, *Dirichlet Latent Variable Model: A Dynamic Model Based on Dirichlet Prior for Audio Processing*, **IEEE Transaction on speech and language processing (2019)** **[Paper]**
  Extended below model to incorporate multi-order bidirectional dependency and to newer applications
- **A. Kumar**, T. Guha, P. Ghosh, *A Dynamic Latent Variable Model for Source Separation*, Int. Conf. on Acoustics, Speech and Signal Processing (**ICASSP'18**) **[Paper]**
  Developed a dynamic latent variable model for time-varying non-negative data. Proposed a novel prior distribution which is particularly suitable for dynamic non-negative data and yields elegant update equations. It also lead to connecting our model to the two popular latent basis learning methods - PLCA and NMF
- **A.Kumar**$^+$, L.Pandey$^+$, V.Namboodiri, *Monoaural audio source separation using VAE*, **Interspeech** '18 **[Paper]**
  Proposed a principled generative approach using VAE for audio source separation. VAE computes efficient Bayesian inference which leads to a continuous latent representation characterizing each source. Our method performed better than best of the relevant methods with 2 dB improvement in the source to distortion ratio
- **A. Kumar**$^+$, K. Morabia$^+$, W. Wang, K. Chang, *VAMWOD: Visual Attention-based Model for Webpage Object Detection*, (Submitted to **CVPR '21**)

SKILLS

- **Programming:** Python, Matlab, C, C++, LaTex, R, JavaScript
- **Other Technologies:** PyTorch, TensorFlow, Caffe, Scikit-learn, AWS, GIT

RELEVANT COURSEWORK

Machine Learning(ML) • ML for Computational Biology • Advanced Biochemistry • Advanced Bioinformatics • Computational Bio-engineering • Computer Vision • Natural Language Processing • Learning Theory • Convex Optimization • Algorithms

PROJECT HIGHLIGHTS

Latent representation based gene regulatory network inference from multi-omic data Aug '19 - Present
Prof. Jian Peng, UIUC [Report]
- Developed a DNN architecture to characterize a gene with complex 3D structure and epigenomic features.
- Implemented a gene interaction ranking framework from ENCODE epigenomic fetaures and L1000 gene expression

Visual grain quality estimation Aug '16 - Nov '16
Prof. T. Guha, IITK, (Best project in Image Processing course) [Report]
- Developed a proof-of-concept for quality estimation of grain from image of a grain sample
- Created a labeled dataset with help of traders and farmers
- Proposed a two-level segmentation method to segment overlapped grains which were further classified as grains / impurities demonstrating performance of proposed technique

Diverse feature selection in Latent Variable models Jan '16 - Apr '16
Prof. P. Rai, IITK [Project Material]
- Reformulated maximum likelihood objective function of Probabilistic PCA and Matrix factorization as maximum aposteriori estimation. Used *determinantal point processes* and *mutual angular regualizer* as priors for diverse selection of latent features.

Unsupervised speaker diarization Jan '16 - Apr '16
Prof. T. Guha, IITK [Project Material]
- Designed an end-to-end pipeline for speaker diarization. The preprocessing stage included extracting MFCC features followed by speaker activity detection. Each speaker was modeled with a multivariate normal distribution. Hypothesis testing with bayesian information crieria was employed for speaker segmentation. Finally clustering was employed to cluster each segments.

Multiple Word Vector Embedding for polysemous words Sep '15 - Nov '15
Prof. A. Mukherjee, IITK [Project Material]
- Extended word vector model to have multiple representations for polysemous words
- Proposed two parametric and two non-parametric solutions that beat state-of-the-art in several specific cases
- Our primary contribution was in developing a non-parametric approach to clustering for optimal number of senses

Visual Odometry in Self-Driving Car Sep '15 - Nov '15
Prof. G. Pandey, IITK [Project Material]
- Developed and implemented visual odometry in self-driving car using opencv library in C++
- SIFT and FAST features were extracted in each frames
- KLT tracker with RANSAC for outlier rejection was employed for tracking features

Multiple Kernel Learning (MKL) Sep '15 - Nov '15
Prof. H. Karnick, IITK [Project Material]
- Implemented and performed experiments to show that MKL automatically learns efficient weighted distribution of multiple kernels and have the potential to handle data coming from heterogeneous sources