

Bodhisattwa Prasad Majumder

Resume

Bangalore, India

(+91) 983 047 1053

✉ bodhisattwapm2017@email.iimcal.ac.in

🌐 www.majumderb.com

Research Interests

Computational Linguistics, Natural Language Processing, Computational Social Science, Complex Networks, Machine Learning, Statistical Modeling [[Research Statement](#)]

Experience

June 2017 – **Research Engineer, Machine Learning, Walmart Labs.**

- Present ○ **Neuro-attribute tagging** using deep recurrent structures to improve faceted product search, LSTM-CRF (with or without attention mechanism) based Sequence to Sequence learning
- **Conversational Agent** with entity extraction and resolution; Guided feedback based learning to improve chat experience starting with very less training data; retrieval based conversations, stateful conversations
- **Attention based OCR** for text extraction from images to improve product catalogs; ML pipeline for text detection, text recognition, information extraction via entity recognition

Education

2015 – 2017 **Master's Degree, Indian Institute of Technology, Kharagpur.**

Major in Data Science & Machine Learning; One semester coursework at Indian Statistical Institute

CGPA: 9.71/10; Rank — 1 out of 51

Master's Thesis: Decoding Consumer Behavior in Retail via Statistical and Deep Learning Frameworks; Unsupervised Determination of Competitive Interrelationships of Products (Walmart Labs)

Advisor: Dr. Animesh Mukherjee, CSE, Indian Institute of Technology Kharagpur

Courses: Machine Learning, Algorithms, Complex Network, Information Retrieval, Data Mining, Multivariate Analysis, Linear Algebra, Markov Chains, Probability and Inference

2011 – 2015 **Bachelor of Engineering, Electronics & Telecomm. Engineering, Jadavpur University.**

CGPA: 9.48/10; Rank — 5 out of 70

Bachelor Thesis: An interactive GUI based Real-time pulse processing system - an on-line streaming data analysis framework (Variable Energy Cyclotron Center, Bhaba Atomic Research Center)

Courses: Data Structures, Operating Systems, Compiler Design & Automata Theory, Signal Processing, Linear Algebra, Differential Equations

Publications

- Amrith Krishna, *Bodhisattwa P. Majumder*, Pawan Goyal, “**An ‘Eklavya’ approach to learning Context Free Grammar rules for Sanskrit using Adaptor Grammar**”, *17th World Sanskrit Conference, 2018 (accepted for presentation)* [[pdf](#)]
- *Bodhisattwa P. Majumder**, Aditya Subramanian*, Abhinandan Krishnan, Shreyansh Gandhi, Ajinkya More, “**Deep Recurrent Neural Networks for Product Attribute Extraction in eCommerce**”, *Submitted in European Conference of Information Retrieval (ECIR), 2018 (* denotes equal contributions)* [[pdf](#)]
- *Bodhisattwa P. Majumder*, Amrith Krishna, Unni Krishnan, Anil K Boga, Animesh Mukherjee, “**What’s in a ‘Meme’? Understanding the Dynamics of Image Macros in Social Media**”, *Preprint manuscript, 2018* [[pdf](#)]
- Denise M. Case, M. Nazif Faqiry, *Bodhisattwa P. Majumder*, Sanjoy Das, and Scott A. DeLoach, “**Implementation of a Two-tier Double Auction for On-line Power Purchasing in the Simulation of a Distributed Intelligent Cyber-Physical System**”, *Advances in Artificial Intelligence, Research in Computer Science, MICAI, 2014* [[pdf](#)]
- *Bodhisattwa P. Majumder*, M. Nazif Faqiry, Sanjoy Das, Anil Pahwa, “**An Efficient Iterative Double Auction for Energy Trading in Microgrids**”, *IEEE Symp. on Computational Intelligence Applications in Smart Grid, 2014* [[pdf](#)]
- *Bodhisattwa P. Majumder*, Ayan Sengupta, Sajal Jain, Parikshit Bhaduri, “**Fault Detection Engine in Intelligent Predictive Analytics Platform in DCIM**”, *Intl. Conf. on Business Analytics & Intelligence, IISc Bangalore, 2016* [[pdf](#)]
- Satrajit Mukherjee, Kunal Pal, *Bodhisattwa P. Majumder*, Chiranjib Saha, B. K. Panigrahi, “**Differential Evolution based Score Level Fusion for Multimodal Biometric Systems**”, *IEEE Symposium on Computational Intelligence in Biometrics and Identity Management, 2014* [[pdf](#)]

- Satrajit Mukherjee, *Bodhisattwa P. Majumder*, Aritran Piplai, Swagatam Das, “**A Novel Fuzzy Non-homogeneity Measure based Kernelised Image Segmentation for Noisy Images**”, *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2014* [pdf]

Book

Applied NLP: Natural Language Processing by Examples, Packt Publishing, UK, [Contents].

Co-authors: Harshit Surana (CMU), Anuj Gupta (IIT-Delhi)

Technical Advisors: Mridul Gupta (Apple, CMU), Kesavan Sivaraman (Amazon, UT Austin)

Scheduled to be published on March 2018

Honors & Awards

- Perfect **SGPA 10** holder in IIT Kharagpur; **Ranked first** in all the semesters in Master's Study
- Endowed with **cash award** for academic excellence (obtaining SGPA > 9.5), Indian Statistical Institute
- **Finalist, Data Science Game '16**, Paris; Represented India (1 out of 3 teams), International Rank 14
- Officially entitled as **contributor** in **NSF-CPS** project (**CNS -1136040**) by PIs, Kansas State University
- **4-year scholarship** for academic excellence, Ministry of Human Resource & Development, India, 2011-15
- **Program Committee** Member, Walmart Labs Data Science Conference, January 2018

Academic Research Experiences

July 2017 – **Computer Science Dept., University of California, San Diego.**

Present **Advisor: Dr. Julian McAuley**

- Distributed behavioral semantic representation of products from large-scale transaction data
- **Temporal Product Embeddings** – We build a temporal binning model to capture the temporal regularities in the product embeddings. Products ‘move’ across the low-dimensional manifold over time. The temporal distribution of spread between product pair indicate the behavioral trend of customer purchase. Spatially similar products across time can be helpful in time-based recommendations, identifying substitutes and complements.

July 2017 – **CNeRG, Computer Science Dept., Indian Institute of Technology Kharagpur.**

Present **Advisor: Dr. Pawan Goyal**

- **Grammar Induction from Morphologically-rich Languages** – We seek to analyze the feasibility of Adaptor Grammar for syntactic analysis and dependency parsing at sentence level using various morphological tags. Apart from Adaptor Grammar's efficacy in extracting sub-word patterns which have been proven to be effective in various supervised tasks, we also have confirmed the feasibility of using Adaptor Grammar for dependency parsing for Sanskrit language.
- Effectiveness of using **Adaptor Grammar** for various **supervised** tasks – compound type classification, structured prediction, identifying the source & derived words for derivational nouns for Sanskrit language
- **Structured Prediction & Syntactic Analysis** using PCFGs – *Poetry to Prose* conversion, Dependency parsing

Jan 2016 – **CNeRG, Computer Science Dept., Indian Institute of Technology Kharagpur.**

Present **Advisor: Dr. Animesh Mukherjee**

- **Diachronic Study of Image Memes** via emerging conventions in a closed community in *Facebook* – We capture the temporal usage pattern to understand how different local subculture affects the image meme generation. We observed the competition between *global* and *local* subculture driven memes as emerging conventions, where *local* subculture enjoys more user engagement than *global* culture.
- **Motif identification** from temporal usage history – We capture the interaction between memes across days within the community using HDP-HMM, a non-parametric Bayesian variation of the infinite HMM. We group the observed days where the states encode possible ‘moods’ on the observed days. We also establish the significance of content-based *familiarity* vs. *freshness* which is key to the growth and evolution of the community.
- Study of **linguistic properties** of memes, modeling sub-patterns, motifs using Adaptor Grammar

Summer 2014 **Electrical and Computer Engineering, Kansas State University.**

Advisor: Dr. Sanjoy Das

- Study of **Multi-Agent Systems (MAS)**, Computational **Mechanism Design**, Multi-tier Double Auctions
- Extension of Kelly's Mechanism in an Iterative Double Auction framework to capture the asymmetric preferences
- Two-tier double auction on distributed devices enabled by Advanced Message Queuing Protocol

Winter 2013 **Biometric Laboratory, Electrical Engineering, Indian Institute of Technology Delhi.**

Advisor: Dr. B. K. Panigrahi

- Multimodal score fusion via Gaussian Kernel smoothing & Differential Evolution for Biometrics
- Modeling of Multiple Member Cluster Constraints based Hierarchical Clustering in large networks

Spring 2016 **Data Science Laboratory**, *Indian Institute of Technology Kharagpur*.

- **Opinion Mining on reviews**: Latent Aspect Rating Regression to capture the latent emphasis on various aspects using generative language model, Stanford NLP Parser & SentiWords based sentiment modeling for aspects
- **LexRank** based summarizer to identify most salient opinion(s) for various aspects

Industry Research Internships

Nov 2016 – **Machine Learning & Data Science Group**, *Walmart Labs*.

- April 2017
- Identifying the competitive interrelationships between products; capturing the idea of like variety seeking versus similarity via latent signal hidden in transaction patterns, Understanding the phenomena of product substitution
 - Unsupervised model for grouping product pairs based on their competitive behavior in Traditional substitutes, Variety substitutes and non-substitutes from product attributes, demographics and consumption history
 - Accepted for *Oral presentation* at Business Analytics and Intelligence conference, 2017, Bangalore

Summer 2016 **R&D Unit**, *GreenField Software*.

- Graph Clustering based framework for correlated alarm system in Data Centers; Hazard function based Markov process to model stochastic recovery behavior

Competitions

Finalist, 2016, *Data Science Game (Microsoft, Capgemini and AxaLabs)*.

- Final, Paris — International Rank **14** out of 140; Prediction of potential insurance buyers from click streams, buyer-buyer network, buyer-seller network metrics; modeling of extreme class-imbalanced data
- 1st Round, Kaggle — Intl. Rank **22** out of 140; Classification of Rooftop orientation from satellite images using ensemble of various CNN architectures including VGG, Inception and LeNet through transfer learning technique

Winner, 2017, *GE Healthcare — GE HealthHack: Visual Content Analysis*.

- 'DeepTrack' – A real time multitenanted system to track patient's movements, expressions and derive insights about patient's current status from streaming videos of ICU environments
- Integrated Face detection, Emotion Identification, Face Tracking modules based on Deep Learning

Winner, 2017, *SyneHack, Synechron*.

- 'DeeMIS' — A system for 'Deep Multimodal Interpretation of Sentiment'
- Deep Networks (CNN-LSTMs) to analyze the emotional content from images, text and audio signals

Patents

- Aditya Subramanian, *Bodhisattwa P. Majumder*, Shreyansh Gandhi, Abhinandan Krishnan, Ajinkya More, "**System and Method for Product Attribute Extraction Using a Deep Recurrent Neural Network**", US Patent, 4661US01, Provisionally filed on 29th October 2017
- Subhasish Misra, Arunita Das, Amlan Das, *Bodhisattwa P. Majumder*, "**Analytical Determination of Competitive Interrelationship between Item Pairs**", US Patent, 4431US01, Filed on 31st October 2017

Program Committee and Invited Talks

Committee Member, Walmart Labs Data Science Conference, *Walmart Labs*, January, 2018.

- Part of the Program Committee, with responsibility of finalizing paper format, reviewer allocation

Member, MUST Research Club, *Microsoft Research, India*, April 2017 - Present.

Talk at Machine Learning Series, *GE Healthcare*, 2018.

- **Topic**: Spanning 3 weeks of talks on Machine Learning and its applications

Talk at Anthill Inside Meetup, *Walmart Labs*, 2017.

- **Topic**: Information Extraction from Images - Application in e-Commerce & Retail

Talk at Graduate Course: Computing for Data Sciences, *Indian Statistical Institute*, 2016.

- **Topic**: Neural Network & Deep Learning: in light of Optimization and Regularization

Skillsets

- **Programming Languages**: Python, R, C++, C, MATLAB, Java
- **Frameworks/Tools/Databases**: Tensorflow, Theano, PyTorch, Keras, PySpark, NLTK, OpenCV, Octave, Teradata, MongoDB, Hive, SQL, $\LaTeX 2_{\epsilon}$; **Web development**: HTML, CSS, JavaScript

Extra Curricular

- 3 photographs selected in International Gallery of B&W Street Photography, *BlackandWhiteStreet.com*
- Co-founder, *ABC Photography*; photographs featured in national newspapers