

Hitesh Golchha | Curriculum Vitae

Volga Complex, 3rd Floor, H.B. Road, Guwahati, India- 781001

☎ +91 70025 41359 • ☎ +91 94350 80504 • ✉ hitesh.iitp@gmail.com
🌐 www.github.com/hitzkrieg

Experience

- **Flipkart** **Bangalore**
Software Development Engineer (Data Science), *Apr 2019–present*
- **AI-NLP-ML Lab** **IIT Patna**
Junior Research Fellow, *Aug–Dec 2018*
- **NLP Lab - Bar Ilan University, Israel** **Ramat Gan, Israel**
Research Internship, *May–July 2017*
- **AI-NLP-ML Lab** **IIT Patna**
Undergraduate Research, *May 2016–June 2018*

Education

- **Indian Institute of Technology Patna** **Patna**
B.Tech. Computer Science and Engineering (CGPA: 8.56), *July 2014– May 2018*

Research and Development

- **NLU for Flipkart App's Bilingual Voice Bot** **Flipkart**
SDE - Data Science *April 2019 - present*
 - Entity extraction: Fine-tuning Transformer Encoders pre-trained on domain and language specific data, CharCNN based models
 - Multi-intent handling: Trained and integrated ML model for segmenting sentences on the presence of multiple intents or absence of punctuation.
 - Data preparation: For pre-training BERT, for spelling variants, plural variants, substitution code for data augmentation
 - Experiments on improving Speech to Text output using re-scoring of beams, specifically for short audio lengths.
- **Style Transfer on Customer Care (NAACL'19)** **IIT Patna**
Junior Research Fellow *under Dr. Asif Ekbal and Prof. Pushpak Bhattacharyya Oct 18 - Dec 18*

- **Courteously Yours: Inducing courteous behavior in Customer Care responses using Reinforced Pointer Generator Network**
- We develop a natural language generation model to transform generic responses to courteous while being contextually coherent.
- Attention based seq2seq model, with Pointer Generator is used. Decoder is conditioned on conversation history (hierarchical LSTMs)
- The developed model is trained using Reinforcement Learning with specifically designed rewards
- Large scale Parallel data is prepared by filtering courteous sentences/expressions from real conversations on Twitter handle of companies.

Multi-Task Learning for NLU (under review)

IIT Patna

○ **Research Assistant** under Dr. Asif Ekbal and Prof. Pushpak Bhattacharyya Aug-Sep 2018

- We develop a Multi-task model to identify the intent, dialog act and slot values for a Natural Language Understanding model on three benchmark datasets.
- We experiment with Character embeddings, self-attention, linguistic embeddings, and different encoders to come up with the final model.
- We also experiment with alternating training methods for multi-domain training

Customer-to-Customer Suggestion Mining (ICON'18)

IIT Patna

○ **B. Tech. Thesis** advised by Prof. Pushpak Bhattacharyya and Dr. Asif Ekbal Aug 2017 - May 2018

- Accepted at *15th International Conference on Natural Language Processing* as a long paper
- Objective is to extract customer to customer suggestions from online reviews
- We develop a semi-supervised learning approach using self-training over a Deep learning based classifier which uses features extracted from three components:
 - *Sequential Encoder* based on LSTM and Inner attention
 - *Convolutional Encoder* using 1-D CNN filters
 - *Linguistic knowledge* extracted from task specific heuristic features and features extracted from dependency parse trees
- Contributions include the use of semi-supervised learning for this problem definition and outperforming the current state of the art techniques.
- Future work includes extending the system to more robust Semi supervised learning algorithms.

Proposition coreference resolution for OKR

Bar Ilan University, Israel

○ **Research Internship** advised by Prof. Ido Dagan

May 2017 – July 2017

- OKR: A Consolidated Open Knowledge Representation for Multiple Texts
- Project consisted of a semantic pipeline generating natural language knowledge graphs by consolidating extracted information from multiple unstructured texts.
- I did the Literature survey, and improved the existing baseline Proposition Coreference resolution system (improvement CONLL 0.56 to 0.65)
- Developed an online clustering algorithm which added new mentions based of pairwise comparison of propositions using fuzzy string matching, wordnet overlap between words and presence of entity coreference between arguments.
- I also integrated generated Natural Language graphs to Neo4J for queries and visualisation

Question Answering from Wikipedia articles

IIT Patna

○ **advised by Dr. Asif Ekbal**

August 2016 - April 2017

It was my first research project. We worked on the WikiQA dataset, aiming to find the answer spans from given relevant Wikipedia articles. We implemented an architecture which used a multi layer perception over the question and potential answer features encoded by LSTM network.

Technical skills

- **Programming Languages:** Proficient in: Python, Java, C. Also proficient in Python Libraries : Numpy, Scikit-Learn, Matplotlib, Tensorflow, Keras, Spacy. Familiar with C++, Matlab, HTML, CSS, PHP, SQL. Also familiar with python Libraries: OpenCV, PyTorch.

Interests and extra-curricular activity

- **Cricket:**
I have been playing cricket since I was 13 years of age, and have got the opportunity to represent Don Bosco School Guwahati and IIT Patna in several tournaments.
- **Dramatics, Anchoring, Technical Clubs:**
I have been a member of a few societies in my undergraduate years, which include Dramatics, Robotics and NJACK(Not just any Computer Club). I have also had the opportunity to anchor different cultural events in IIT Patna, which include Asian Conference on Solid State Ionics-2016, Anwesha-2016, and Reverberance-2015.
- **Technical Co-ordinator, Anwesha:**
I have also had the privilege of being the Technical Co-ordinator of Anwesha, the annual Techno-Cultural Intercollegiate Festival of IIT Patna organised by the student fraternity.