

IrshadAhmad Bhat | CurriculumVitae

Khankahbagh 129, 192121 J & K – India

📞 (+91) 7893940615 • ✉ bhatirshad127@gmail.com

📄 (Google Scholar) • 📄 Github

Work Experience

ML/NLP Research Scientist

Feb 2018–March 2019

Active Intelligence Pte Ltd (Active.ai)

Bangalore, India

○ Tasks:

- **Neural Machine Translation (NMT)**: Created and deployed NMT models (based on transformer, convolutional and BiLSTM neural networks) for a wide range of languages like Arabic, Spanish, Vietnamese, French, Hindi, Tamil, Bengali etc.
- **Intent Classifiers for Conversational Data**: Worked with Bert-based models to deploy accurate Intent classifiers with zero pre/post processing of data.
- **Sentiment Analysis**: BERT-based sentiment classification of conversational data.
- **Fine-tuning APIs**: Deployed easy to use fine-tuning APIs for all of the above models.

Machine Learning Researcher

May 2019–Dec. 2020

IMS, University of Stuttgart

Stuttgart, Germany

Worked on the project "Misunderstandings in/from Text (MisT)". The goal of the project is to create resources and methods to prevent these misunderstandings. Towards this goal I created a dataset *wikiHowToImprove: A resource on edits in wikiHow articles*. (published in LREC 2020) and provided benchmark models for distinguishing older versions of text from newer versions and predicting revision requirements (published in EMNLP 2020).

Education

Master's in Computer Science and Engineering

Jan. 2015–Aug. 2018

IIT-H Gachibowli

Hyderabad, India

Bachelor's in Computer Science and Engineering

Sept. 2010–Aug. 2014

Govt. College of Engineering and Technology

Jammu, India

MS Thesis

Title: *Universal Dependency Parsing of Hindi-English Code-switching*

Supervisor: Dr. Manish Shrivastava

Research Interests

Code Switching and BERT-like multilingual models.: Application of BERT like models (BERT, RoBERTa, XLNET, GPT-2, ELECTRA etc.) for improving language processing of code switching texts like Hinglish.

Syntactic Parsing: Application of fast and efficient transition-based/attention-based parsing algorithms for syntactic analysis of natural language texts.

Machine Translation: Neural Machine Translation using Transformer, Convolutional and BiLSTM networks.

Machine Transliteration: Bidirectional script transliteration between Indic and Roman scripts.

ML/NLP: Application of different ML/NLP for tasks like classification, tagging, entailment, machine comprehension, parsing etc.

Language Resource Contributions

2014–present

- **wikiHowToImprove:**

Created a dataset of revised and unrevised sentences from revision histories of wikiHow articles. The dataset can be used for various tasks like predicting revision requirements, grammar correction, ambiguity detection, machine translation refinement etc.

- **Hindi-Urdu Treebanking Project:**

As an RA for Hindi-Urdu treebanking project, I have been maintaining the annotation pipeline of the Urdu dependency treebank. The work mainly involves creating and assigning annotation tasks to the annotators.

- **Universal Dependencies:**

My contributions include addition of datasets with Universal Dependencies for Code-Switching Hindi-English and movie scripts.

- **Google Summer of Code 2016 (GSoC):**

I worked for Libindic organization under the Google Summer of Code 2016 and contibuted towards auto-matic script transliteration between scheduled languages of India including English.

Teaching Experience

Teaching Assistant at IIT Hyderabad

- **2015:** STATISTICAL METHODS IN AI (CSE471)

Technical Qualifications

Programming: PYTHON, C, C++, PERL, SHELL SCRIPTING, JAVA SCRIPT

Deep Learning Frameworks: PYTORCH, TENSORFLOW, KERAS, DYNET

Operating Systems: LINUX, WINDOWS

Protocols: L^AT_EX, HTML, XML, JSON, CSS

Editors: VIM, NANO, EMACS

Publications

Conferences Papers

2020

- IRSHAD AHMAD BHAT and Michael Roth. *Towards Modeling Revision Requirements in wikiHow Instructions*. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020).
- Talita Anthonia, IRSHAD AHMAD BHAT, *wikihowtoimprove: A resource and analyses on edits in instructional texts*. Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020).

2018

- IRSHAD AHMAD BHAT, Riyaz Ahmad Bhat and Manish Shrivastava, Dipti Misra Sharma. *Universal Dependency Parsing for Hindi-English Code-switching*. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2018),
- Riyaz Ahmad Bhat, IRSHAD AHMAD BHAT and Srinivas Bangalore *The SLT-Interactions Parsing System at the CoNLL 2018 Shared Task*. Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies.

2017

- Riyaz Ahmad Bhat, IRSHAD AHMAD BHAT and Dipti Misra Sharma. *Leveraging Newswire Treebanks for Parsing Conversational Data with Argument Scrambling*. In Proceedings of the 15th International Conference on Parsing Technologies (IWPT) 2017, Pisa, Italy.
- IRSHAD AHMAD BHAT, Riyaz Ahmad Bhat, Manish Shrivastava and Dipti Misra Sharma. *Joining Hands: Exploiting Monolingual Treebanks for Parsing of Code-mixing Data*. In Proceedings of the European Chapter of the Association of Computational Linguistics (EACL) 2017, Valencia, Spain.

2016

- Riyaz Ahmad Bhat, IRSHAD AHMAD BHAT and Dipti Misra Sharma. *A House United: Bridging the Script and Lexical Barrier between Hindi and Urdu*. The 26th International Conference on Computational Linguistics (COLING 2016).

2015

- Riyaz Ahmad Bhat, IRSHAD AHMAD BHAT and Dipti Misra Sharma. *Improving Dependency Parsing of Hindi and Urdu by Modeling Syntactically Relevant Phenomena*. ACM Transactions on Asian and Low-Resource Language Information Processing (TALIP).

2014

- IRSHAD AHMAD BHAT, Vandan Mujadia, Aniruddha Tammewar, Riyaz Ahmad Bhat and Manish Shrivastava. IIT-H SYSTEM SUBMISSION FOR FIRE2014 SHARED TASK ON TRANSLITERATED SEARCH. The Forum for Information Retrieval Evaluation (FIRE 2014).