Abhijit Mishra

Research Scientist
IBM Research, Bangalore, India
+918879006045
abhijimi@in.ibm.com
https://abhijitmishra.github.io

ABOUT ME

I am currently a part of *IBM Research*, Bangalore, India, serving as *Research Scientist* in the division of *AI-Tech*. I am involved in multiple projects based on Natural Language Generation (NLG), *viz.*, (1) Unsupervised Controllable Language Transformation, and (2) Structured Data Summarization.

Prior to joining IBM Research, I was a Ph.D. student in the Department of Computer Science and Engineering, Indian Institute of Technology Bombay (graduated in 2017). I was advised by Prof. Pushpak Bhattacharyya for my research on *Cognitive NLP* which aims to (i) uncover the cognitive underpinnings of human language processing and (ii) translate the insights into better language processing systems. For this purpose, *Eye-Tracking* technology was employed to record and analyze the eye movement patterns of human annotators for various language processing tasks such as Translation, Sentiment and Sarcasm Analysis.

Apart from thesis related work, I contributed to projects related to Machine Translation, Social Media Text Analysis, Crowdsourcing, and Development of Resources and Tools for Indian Language Processing.

APPOINTMENTS

Research Scientist Mar 2017-
IBM Research AI, Bangalore, India

- Department: AI Tech
- Current Focus: Systems and services involving Natural Language Generation from Structured Data, Unsupervised Text Simplification, Controllable Text Generation

Research Assistant July 2011-Mar 2017
IIT Bombay, Mumbai, India

- Project 1: *Cognitive NLP* - Harnessing Cognitive Information for Natural Language Processing, An Investigation based on Eye-tracking (Thesis Project)
- Project 2: *Shata-Anuvaadak (100 Translators)* - A Multilingual Machine Translation Framework for Indian Languages
- Project 3: Generation of Multilingual Resources for Indian Language Processing through Crowdsourcing
- Project 4: A Computational Approach to Automatic Prediction of Drunk Texting

Visiting Researcher June 2012-Aug 2012
Copenhagen Business School, Denmark

- *Research Topic:* Eye Tracking Applications in Translation Process Research
- Analyzed translators' activities collected in the form of eye-movement and keystroke patterns

- Correlated complexity of eye movement patterns, believed to be an indicator of perceived complexity, with linguistic subtleties of the translated text

Assistant System Engineer August 2010-June 2011
Tata Consultancy Services Limited, Bangalore India

- *Role:* Developer, *Client:* Allianz-USA
- Developed and supported a module for automatically managing documents using Java, VB Script, HTML-CSS

EDUCATION

Ph.D. Computer Science and Engineering 2011-2017
Indian Institute of Technology Bombay, Mumbai, India
"Harnessing Cognitive Information for Natural Language Processing: An Investigation based on Eye-tracking"
Thesis Advisor: Pushpak Bhattacharyya, Professor, Dept. of C.S.E, IIT Bombay

B.Tech. Computer Science and Engineering 2006-2010
College of Engineering and Technology, Bhubaneswar, India
CPI 9.03 (out of 10)

BOOK PUBLISHED

- Abhijit Mishra and Pushpak Bhattacharyya, **"Cognitively Inspired Natural Language Processing"**, Springer, Singapore. DOI: https://doi.org/10.1007/978-981-13-1516-9 . URL: https://link.springer.com/book/10.1007/978-981-13-1516-9#about

PUBLICATIONS

Main Conference

18. Anirban Laha, Parag Jain, Abhijit Mishra, Karthik Sankaranarayanan. 2019. Scalable Micro-planned Generation of Discourse from Structured Data. arXiv preprint arXiv:1810.02889. Under review at Computational Linguistics, MIT Press

17. Sai Surya, Abhijit Mishra, Anirban Laha, Parag Jain, Karthik Sankaranarayanan. 2019. Unsupervised Neural Text Simplification. arXiv preprint arXiv:1810.07931.

16. Parag Jain, Abhijit Mishra, Amar P. Azad, Karthik Sankaranarayanan. 2019. *Unsupervised Controllable Text Formalization.* **AAAI 2019**, Hawaii, USA, 27th Jan - 1st Feb, 2019

15. Sandeep Mathias, Diptesh Kanojia, Kevin Patel, Samarth Agrawal, Abhijit Mishra and Pushpak Bhattacharyya. 2018. *Eyes are the Windows to the Soul: Predicting the Rating of Text Quality Using Gaze Behaviour.* **ACL 2018**, Melbourne, Australia, 15-20 July, 2018.

14. Vitobha Munigala, Abhijit Mishra, Srikanth Govindaraj Tamilselvam, Shreya Khare, Riddhiman Dasgupta and Anush Sankaran. 2018. *PersuaAIDE ! An Adaptive Persuasive Text Generation System for Fashion Domain.* **WWW 2018**, Lyon, France, 23th April - 27th April, 2018

13. Abhijit Mishra, Srikanth Tamilselvam, Riddhiman Dasgupta, Seema Nagar and Kuntal Dey. 2018. *Cognition-Cognizant Sentiment Analysis with Multitask Subjectivity Summarization based on Annotators' Gaze Behavior.* **AAAI 2018**, New Orleans, USA, 2nd February - 7th February, 2018

12. Srikanth Tamilselvam, Seema Nagar, Abhijit Mishra and Kuntal Dey. 2017. *Graph Based Sentiment Aggregation using ConceptNet Ontology.* **IJCNLP 2017**, Taipei, Taiwan, 27 November-1st December, 2017

11. Shweta Garg, Sudhanshu S Singh, Abhijit Mishra and Kuntal Dey. 2017. *CVBed: Structuring CVs using Word Embeddings.* **IJCNLP 2017**, Taipei, Taiwan, 27 November-1st December, 2017

10. Joe Cheri Ross, Abhijit Mishra, Kaustuv Kanti Ganguli and Pushpak Bhattacharyya. 2017. *Identifying Raga Similarity Through Embeddings Learned from Compositions' Notation.* **ISMIR 2017**, Suzhou, China, 23-28 October, 2017

9. Abhijit Mishra, Kuntal Dey and Pushpak Bhattacharyya. 2017. *Learning Cognitive Features from Gaze Data for Sentiment and Sarcasm Classification using Convolutional Neural Network.* **ACL 2017**, Vancouver, Canada, 30 July-4 August, 2017

8. Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, Pushpak Bhattacharyya. 2017. *Scanpath Complexity: Modeling Reading Effort using Gaze Information.* **AAAI 2017**, San Francisco, USA, 4-9 February, 2017

7. Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey and Pushpak Bhattacharyya. 2016. *Harnessing Cognitive Features for Sarcasm Detection.* **ACL, 2016**, Berlin, Germany, 7-12 August, 2016

6. Abhijit Mishra, Diptesh Kanojia, Kuntal Dey, Seema Nagar and Pushpak Bhattacharyya. 2016. *Leveraging Cognitive Features for Sentiment Analysis.* **CoNLL 2016**, Berlin, Germany, August 11-12, 2016

5. Abhijit Mishra, Diptesh Kanojia and Pushpak Bhattacharyya. 2016. *Predicting Readers' Sarcasm Understandability by Modelling Gaze Behaviour.* **AAAI 2016**, Phoenix, USA, Feb 12-17, 2016

4. Aditya Joshi, Abhijit Mishra, Balamurali AR, Pushpak Bhattacharyya, Mark J Carman. 2015. *A Computational Approach for Automatic Prediction of Drunk-texting.* **ACL 2015**, Beijing, China, July 2015

3. Aditya Joshi, Abhijit Mishra, Nivvedan Senthamilselvan and Pushpak Bhattacharyya. 2014. *Measuring Sentiment Annotation Complexity of Text.* **ACL 2014**, Baltimore, USA, 23-25 June, 2014

2. Anoop Kunchukuttan, Abhijit Mishra, Rajen Chatterjee, Ritesh Shah and Pushpak Bhattacharyya. 2014. *Shata-Anuvadak: Tackling Multiway Translation of Indian Languages.* **LREC 2014**, Rekjyavik, Iceland, 26-31 May, 2014

1. Abhijit Mishra and Pushpak Bhattacharyya. 2013. *Automatically Predicting Sentence Translation Difficulty.* **ACL 2013**, Sofia, Bulgaria, 4-9 August, 2013

Workshop and System Papers

8. Parag Jain, Priyanka Agrawal, Abhijit Mishra, Mohak Sukhwani and Anirban Laha. 2017. *Story Generation from Sequence of Independent Short Descriptions.* **ML4Creativity**, SIGKDD Workshop, Halifax, Nova Scotia - Canada, 2017

7. Joe Cheri, Abhijit Mishra and Pushpak Bhattacharyya. 2016. *Leveraging Annotators' Gaze Behaviour for Coreference Resolution.* ACL 2016 Workshop on **Cognitive Aspects of Computational Language Learning** (CogACLL 2016) at ACL 2016, Berlin, Germany, August 11, 2016

6. Diptesh Kanojia, Shehzaad Dhuliawala, Abhijit Mishra, Naman Gupta and Pushpak Bhattacharyya. 2015. *TransChat: Cross-Lingual Instant Messaging for Indian Languages.* **ICON 2015**, December 2015

5. Abhijit Mishra, Aditya Joshi and Pushpak Bhattacharyya. 2014. *A cognitive study of subjectivity extraction in sentiment annotation.*, 5th Workshop on **Computational Approaches to Subjectivity, Sentiment and Social Media Analysis** (WASSA 2014), Baltimore, USA, 27 June, 2014

4. Anoop Kunchukuttan, Ratish Pudupully, Rajen Chatterjee, Abhijit Mishra, Pushpak Bhattacharyya. 2014. *The IIT Bombay SMT System for ICON 2014 Tools Contest.* NLP Tools Contest at **ICON 2014** (ICON 2014), Goa, India, Dec 2014

3. Piyush Dungarwal, Rajen Chatterjee, Abhijit Mishra, Anoop Kunchukuttan, Ritesh Shah and Pushpak Bhattacharyya. 2014. *The IIT Bombay Hindi-English Translation System at WMT 2014.* 9th Workshop on **Statistical Machine Translation** (WMT14), Baltimore, USA, 26-27 June, 2014

2. Anoop Kunchukuttan, Rajen Chatterjee, Shourya Roy, Abhijit Mishra and Pushpak Bhattacharyya. 2013. *TransDoop: A Map-Reduce based Crowdsourced Translation for Complex Domain.* **ACL 2013**, Sofia, Bulgaria, 4-9 August, 2013

1. Abhijit Mishra, Michael Carl and Pushpak Bhattacharyya. 2012. *A Heuristic Based Approach for Systematic Error Correction of Gaze Data for Reading.* First Workshop on **Eye Tracking and NLP**, part of **COLING 2012**. Mumbai, India, 15 Dec, 2012

SIGNIFICANT PROJECTS

1. **Natural Language Generation (NLG) from Structured Data** *(Duration: 2017- , Status: Ongoing)*: The research aims at generating natural language descriptions from structured data such as knowledge graphs, tables *etc.* Motivated by the need to approach this problem in a manner that is scalable and adaptable to newer domain (unlike existing related systems that are rule/template-based or end-to-end neural systems), we introduce scalable modular approaches that do not require any labelled data for generation. Rather, these systems require only large scale unlabelled text and basic NLP tools such as Part of Speech taggers. Our initial experiments on a benchmark mixed domain dataset reveal the superiority of our framework over various existing data-to-text systems. We are currently focusing on generation of interesting narratives from structured data.

2. **Unsupervised Controllable Language Generation** *(Duration: 2017-, Status: Ongoing)*: Like data-to-text NLG, scalable and interpretable solutions are also elusive for text-to-text

NLG problems such as text-simplification, formalization, purpose paraphrasing. We aim to devise new unsupervised learning schemes for text-to-text NLG problems. So far we have proposed novel and practical solutions for unsupervised text simplification and controllable text transformation. We have also attempted to provide solutions for sarcasm generation, a highly nuanced task that requires language understanding at deep semantic and pragmatic levels.

3. **Cognitive NLP through Eye-tracking** *(Duration: 2013-2017, Status: Completed, Remark: PhD. Thesis)*: The research attempts to gain insights into the cognitive underpinnings of human language processing and understanding. The insights are then translated to methods and models that contribute to the field of NLP by achieving the following objectives: (1) Optimizing Human *Annotation Effort* for better annotation management for NLP, and (2) Improving existing NLP systems by introducing cognitive features.

   Today's NLP is highly statistical in nature and needs massive amount of human annotated data. In our setting, apart from collecting the annotations, we aim to record annotators' activities in the form of their eye movement patterns, key-strokes and neuro-eletric signals obtained using EEG. Through a series of studies using eye-tracking alone, we show that data of such kind, can be used to model complexities of tasks like translation and sentiment annotation, where eye-movement data is used to label training data that model annotation effort for the specified tasks. This can be useful for better annotation management (for example, proposing better annotation cost models). We also show that eye movement data can also be used to extract *Cognition Driven* Features, to be used to be used for difficult NLP tasks like Sentiment Analysis and Sarcasm Detection. Our proposed approaches consistently perform better than state-of-the-art sentiment and sarcasm classifiers, showing that cognitive features can be useful for tasks that are nuanced by linguistic subtleties. For more information, visit www.cfilt.iitb.ac.in/cognitive-nlp/

4. **Indian Language Machine Translation** *(Duration: 2013-2014, Status: Completed)*: We developed a compendium of 110 Statistical Machine Translation systems built from parallel corpora of 11 Indian languages belonging to the Indo-Aryan and Dravidian families. We analyze the relationship between translation accuracy and the language families involved. We feel that insights obtained from this analysis will provide guidelines for creating machine translation systems for specific Indian language pairs. For our studies, we built phrase based systems and some extensions. Across multiple languages, we show improvements on the baseline phrase based systems using these extensions: (1) Source side reordering for English-Indian language translation, and (2) Transliteration of untranslated words for Indian language-Indian language translation. These enhancements harness shared characteristics of Indian languages. To stimulate similar innovation widely in the NLP community, we have made the trained models for these language pairs publicly available. The system is available at: http://www.cfilt.iitb.ac.in/indic-translator

5. **Crowdsourcing for NLP Resources** *(Duration: 2011-2013, Status: Completed)*: We developed a framework (Funded by Xerox Research Center, India) that helps an NLP developer to customize and float linguistic annotation tasks through popular crowdsourcing service providers (like Amazon's Mechanical Turk). Though the framework is generic and flexible enough to tackle different linguistic tasks, we took Machine Translation as our use case to demonstrate its efficacy. We show that, using this framework, multilingual translation parallel corpora could be collected and quality controlled with less expenditure in comparison to the traditional way of outsourcing translation tasks to professional translators.

FEDERAL FUNDING

- Cognitive Science Research Initiative Scheme (2013), Department of Science and Technology (Govt. of India), for research on Cognitive Natural Language Processing based on Eye-tracking and EEG (INR 40 Lacs, PI - Prof. Pushpak Bhattacharyya)
- Research Infrastructure Funding Scheme, IIT Bombay (2014) for Cognitive NLP Lab Setup (INR 42 Lacs, PI - Prof. Pushpak Bhattacharyya)

SELECTED PATENTS

5. Abhijit Mishra, Enara C Vijil, Seema Nagar, Kuntal Dey. 2018. Memory Augmented Factoid Question Answering Influenced by Human Gaze Behavior and Text Metadata Generation. *Status: FILED by IBM*, US Patents and Trademarks Office (USPTO)
4. Abhijit Mishra, Parag Jain, Anirban Laha, Karthik Sankaranarayanan. 2018. System and Method for Unsupervised Scalable Generation of Variable Natural Language Descriptions from Structured Data. *Status: FILED by IBM*, US Patents and Trademarks Office (USPTO)
3. Parag Jain, Amar Azad, Abhijit Mishra, Karthik Sankaranarayanan. 2018. System and Method for Unsupervised Tunable Stylized Text Transformation. *Status: FILED by IBM*, US Patents and Trademarks Office (USPTO)
2. Abhijit Mishra, Anirban Laha, Parag Jain, Karthik Sankaranarayanan. 2018. Interactive collaborative document writing system offering real-time natural language evaluations and suggestions. *Status: FILED by IBM*, US Patents and Trademarks Office (USPTO)
1. Abhijit Mishra, Parag Jain, Amar Azad, Karthik Sankaranarayanan. 2018. System and Method for Unsupervised Fuzzily and Crisply Defined Tunable Text Multi-style Transformation. *Status: Rated FILE by IBM*, US Patents and Trademarks Office (USPTO)

TUTORIALS AND INVITED TALKS

- *To be presented at* ***ACL 2019*** *(28 July-2 August):* Storytelling from Structured Data and Knowledge Graphs : An NLG Perspective.
- *18 Jan, 2019:* "Tutorial on Natural Language Generation and its Applications",Indian Institute of Science, Bangalore, India
- *28 July, 2018:* Invited talk at The FAER Faculty Development Workshop on AI&ML, M.S. Ramaiah University titled "Understanding how machines understand us: A perspective on Natural Language Processing", M.S. Ramaiah University, Bangalore, India
- *26 Apr, 2018:* Tutorial on "Cognitively Inspired Natural Language Understanding and Generation", Dharmsinh Desai University , Gujarat, India
- *2 Feb, 2018:* Talk on paper "Cognition-Cognizant Sentiment Analysis with Multitask Subjectivity Summarization based on Annotators' Gaze Behavior", AAAI 2018 Conference, New Orleans, USA

- *20 Jan, 2018:* Tutorial on "Natural Language Generation", Indian Institute of Science, Bangalore, India
- *31 Jul, 2017:* Talk on paper "Learning Cognitive Features from Gaze Data for Sentiment and Sarcasm Classification using Convolutional Neural Network", ACL 2017 Conference, Vancouver, Canada
- *08 Aug, 2016:* Presentation on "Harnessing Cognitive Features for Sarcasm Detection.", ACL 2016 Conference, Berlin, Germany
- *20 June 2016:* Tutorial on "Natural Language Processing and Machine Learning", VIVA Institute of Technology, Mumbai, India
- *21 Jan, 2015:* Talk on "Cognitive NLP", IIT Bombay (Target Audience: Visiting Team, NIST, USA)
- *10 Jul, 2014 - 18 Jul, 2014:* Tutorial on "Natural Language Processing", Samsung Research Lab, Bangalore
- *28 Nov, 2013:* Talk on "Eye Tracking Applications in Translation Process Research", JSS Academy of Science, Noida, India
- *10 Aug, 2013:* Talk on "Estimation of Text Translation Complexity", Copenhagen Business School, Copenhagen, Denmark

AWARDS AND RECOGNITIONS

- Recipient of *IBM Ph.D. Fellowship* award for the academic year of 2015-2016
- Recipient of institute level, department level conference travel grants and Microsoft and IBM Travel Grant to attend ACL (2013, 2016, 2017) and AAAI (2016, 2018) conferences
- Received *Regional Runner-Up Award* for the region India and South-east Asia in the global Windows 7 Coding contest (Code 7 Contest, 2009) organized by Microsoft Corporation for developing a gesture based interactive human computer interface that can run on any Windows based system with a low cost web-camera

ACADEMIC SERVICES

- Program committee member for **EMNLP2019**, **ACL2019**, **EMNLP2018**, **ACL (2018[1],2017)**, **COLING (2018,2016)**, **LREC 2018**, **NAACL 2016**, **WMT (2015,2016,2017)**, **ICON (2014, 2015, 2016, 2017), TALLIP Journal (2018)**, **IEEE Transactions on Affective Computing (2018)**
- Organizing committee member, **COLING 2012**, Mumbai, India
- Co-guided the B.Tech thesis projects of Nivvedan S. (C.S.E., IIT Bombay, 2014), Shubham S. (C.S.E., IIIT Allahabad, 2014), and Krishna G. (C.S.E., IIT Bombay, 2017)
- Teaching Assistant for Natural Language Processing Course (CS 626-460), Autumn, 2014

[1] adjudged as top-reviewer

TECHNICAL SKILLS

- *Programming Languages:* Python, C, C++, Java, Perl, HTML and CSS
- *Language processing and machine learning tools:* PyTorch and Keras for Deep Learning, NLTK, Stanford Core NLP tool, Moses Statistical Machine Translation Toolkit, Weka, Scikit-learn
- *Eye-tracking/EEG devices and tools:* SR research Eyelink 1000, Tobii eye-tracking systems, Translog-II-eye tracking tool (developed some portions of this tool), Biosemi EEG

PERSONAL DETAILS

Date of Birth: 3rd March, 1989
Permanent Address: S/O- Bibhuti Bhusan Mishra, Matiapada, Puri -2, Odisha
Languages Known: English, Hindi, Odia, and German (basic)