Gaurav Kumar

226, Hackerman Hall
3400, N. Charles St.
Baltimore, MD, 21210

(413) 362-2401
gaurav.bison@gmail.com
noisychannel.github.io

Research Machine Translation, Automatic Speech Recognition, Natural Language Processing and Machine Learning.

Education

The Johns Hopkins University 2014-2021
Research Assistant (Center for Language and Speech Processing)
Ph.D. in Computer Science
Research Advisors : Dr. Sanjeev Khudanpur, Dr. Philipp Koehn

The Johns Hopkins University 2012-2014
M.S.E. in Computer Science
Research Advisors : Dr. Sanjeev Khudanpur, Dr. Jason Eisner

Vellore Institute of Technology University, India 2004-2008
B.Tech. in Computer Science and Engineering

Publications

Gaurav Kumar, Philipp Koehn, and Sanjeev Khudanpur, **Learning Feature Weights using Reward Modeling for Denoising Parallel Corpora** in *arXiv preprint*: 2103.06968, 2021 [PDF]

Gaurav Kumar, Philipp Koehn, and Sanjeev Khudanpur, **Learning Policies for Multilingual Training of Neural Machine Translation Systems** in *arXiv preprint*: 2103.06964, 2021 [PDF]

Gaurav Kumar, George Foster, Colin Cherry, Maxim Krikun, **Reinforcement Learning based Curriculum Optimization for Neural Machine Translation** in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL), 2019 [PDF]

Xuan Zhang, Pamela Shapiro, Gaurav Kumar, Paul McNamee, Marine Carpuat, Kevin Duh, **Curriculum Learning for Domain Adaptation in Neural Machine Translation** in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL), 2019 [PDF]

Xuan Zhang, Gaurav Kumar, Huda Khayrallah, Kenton Murray, Jeremy Gwinnup, Marianna J Martindale, Paul McNamee, Kevin Duh, Marine Carpuat, **An Empirical Exploration of Curriculum Learning for Neural Machine Translation** in *arXiv preprint*: 1811.00739, 2018 [PDF]

Huda Khayrallah, Gaurav Kumar, Kevin Duh, Matt Post, Philipp Koehn, **Neural Lattice Search for Domain Adaptation in Machine Translation** in *Proceedings of the Eighth International Joint Conference on Natural Language Processing* (IJCNLP), 2017 [PDF]

Jan Trmal, Gaurav Kumar, Vimal Manohar, Sanjeev Khudanpur, Matt Post, Paul McNamee, **Using of heterogeneous corpora for training of an ASR system**, *arXiv preprint*: 1706.00321, 2017 [PDF]

Graham Neubig, Chris Dyer, Yoav Goldberg, Austin Matthews, Waleed Ammar, Antonios Anastasopoulos, Miguel Ballesteros, David Chiang, Daniel Clothiaux, Trevor Cohn, Kevin Duh, Manaal Faruqui, Cynthia Gan, Dan Garrette, Yangfeng Ji, Ling-

peng Kong, Adhiguna Kuncoro, Gaurav Kumar, Chaitanya Malaviya, Paul Michel, Yusuke Oda, Matthew Richardson, Naomi Saphra, Swabha Swayamdipta, Pengcheng Yin, **DyNet: The Dynamic Neural Network Toolkit**, *arXiv preprint*:1701.03980, 2017 [PDF]

Shuoyang Ding, Huda Khayrallah, Philipp Koehn, Matt Post, Gaurav Kumar, and Kevin Duh, **The JHU Machine Translation Systems for WMT 2017** in *Proceedings of the Second Conference on Machine Translation* (WMT), 2017 [PDF]

Gaurav Kumar, Graeme Blackwood, Jan Trmal, Daniel Povey and Sanjeev Khudanpur, **"A Coarse-Grained Model for Optimal Coupling of ASR and SMT Systems for Speech Translation"** in *Conference on Empirical Methods in Natural Language Processing* (EMNLP), 2015 [PDF]

Matt Post and Yuan Cao and Gaurav Kumar, **"Joshua 6: A phrase-based and hierarchical statistical machine translation system"** in *Prague Bulletin of Mathematical Linguistics*, 2015 [PDF]

Gaurav Kumar, Yuan Cao, Ryan Cotterell, Chris Callison-Burch, Daniel Povey and Sanjeev Khudanpur, **"Translations of the CALLHOME Egyptian Arabic Corpus for Conversational Speech Translation"** in *Proceedings of the International Workshop on Spoken Language Translation* (IWSLT), 2014 [PDF]

Gaurav Kumar, Matt Post, Daniel Povey and Sanjeev Khudanpur, **"Some Insights from Translating Conversational Telephone Speech"** in *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), 2014 [PDF]

Sarana Nutanong, Yanif Ahmad, I-Jeng Wang, Jeliazko Jeliazkov, Gaurav Kumar and Thomas B. Woolf, **Learning about transitions: Adaptive Controls for the Molecular Dynamics Database** in *The 58th Annual Meeting of the Biophysical Society*, 2014

Matt Post, Gaurav Kumar, Adam Lopez, Damianos Karakos, Chris Callison-Burch and Sanjeev Khudanpur, **"Improved Speech-to-Text Translation with the Fisher and Callhome Spanish-English Speech Translation Corpus"** in *Proceedings of the International Workshop on Spoken Language Translation* (IWSLT), 2013 [PDF]

Employment

Research Assistant, **CLSP, Johns Hopkins University**, Baltimore 2013-2021
Worked on problems related to Machine Translation and Speech Recognition.

Software Developer Intern, **Google AI (Translate)**, Montreal 2018-2018
Developed a Reinforcement-learning based technique for noisy data selection for Neural Machine Translation.

Summer Research Intern, **IBM T.J. Watson Research Center**, NY 2014-2014
Worked on research involving optimal couplings for Machine Translation and Automatic Speech Recognition systems.

Lead Architect, **Blisstering Solutions Pvt. Ltd.**, India 2008-2012
Led a team which developed multi-media recommendation engines and web-application frameworks for a variety of clients.

Product Manager, Team Lead **Banyan Mobile Pvt. Ltd., India** 2010-2012
Built recommendation engines and web-application frameworks for telecommunication companies.

CTO, Trustee, **Immunize India Charitable Trust** 2010-present
Responsible for the technology platform which supports the vaccination efforts.

Other Employment / Workshops

Jelinek Summer Workshop on Speech and Language Technology 2017
Neural Machine Translation with Minimum Parallel Resources
Supervisors: Dr. George Foster, Dr. Colin Cherry

Shared Task, Conference on Machine Translation (WMT) 2017
Neural Lattice Search Methods

Third Machine Translation Marathon in the Americas (MTMA) 2017
Low and zero resource Neural Machine Translation

Jelinek Summer Workshop on Speech and Language Technology 2015
Context-aware Neural Machine Translation
Supervisor: Dr. Chris Dyer

Summer Camp for Applied Language Exploration (SCALE) 2015
Speech Translation Methods for Low Resource Languages
Supervisors: Dr. Matt Post, Dr. Sanjeev Khudanpur

First Machine Translation Marathon in the Americas (MTMA) 2015
Explicit Context Encoding with Recurrent Architectures

DARPA Broad Operational Language Translations (BOLT) 2014
Speech Translation for Egyptian-Arabic

Summer Camp for Applied Language Exploration (SCALE) 2013
Improved Speech Translation Architectures
Supervisors: Dr. Matt Post, Dr. Sanjeev Khudanpur

Relevant Open Source Contributions

Kaldi Speech Recognition Toolkit
Joshua Statistical Machine Translation Toolkit
Dynet Neural Network Toolkit

Skills

Programming Languages: Python, Java, Perl, Bash, PHP, C, C++, LaTeX
Deep Learning Frameworks: PyTorch, Tensorflow, MXnet
Deep Learning toolkits: Fairseq, Lingvo, Sockeye, OpenNMT, Theano, Kaldi

References

Dr. Philipp Koehn (phi@jhu.edu), Johns Hopkins University
Dr. Sanjeev Khudanpur (khudanpur@jhu.edu), Johns Hopkins University
Dr. Matt Post (post@cs.jhu.edu), Johns Hopkins University
Dr. Kevin Duh (kevinduh@cs.jhu.edu) Johns Hopkins University

Updated on March 16, 2021.