

CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday – April 5, 2019

Paper Review: Learning Noise-Invariant Representations for Robust Speech Recognition by Davis Liang, Zhiheng Huang, and Zachary C. Lipton
Zhiping Mao & Xuhui Meng

Despite rapid advances in speech recognition, current models remain brittle to superficial perturbations to their inputs. Small amounts of noise can destroy the performance of an otherwise state-of-the-art model. To harden models against background noise, practitioners often perform data augmentation, adding artificially-noised examples to the training set, carrying over the original label. In this paper, we hypothesize that a clean example and its superficially perturbed counterparts shouldn't merely map to the same class — they should map to the same representation. We propose invariant representation-learning (IRL): At each training iteration, for each training example, we sample a noisy counterpart. We then apply a penalty term to coerce matched representations at each layer (above some chosen layer). Our key results, demonstrated on the Librispeech dataset are the following: (i) IRL significantly reduces character error rates (CER) on both 'clean' (3.3% vs 6.5%) and 'other' (11.0% vs 18.1%) test sets; (ii) on several out-of-domain noise settings (different from those seen during training), IRL's benefits are even more pronounced. Careful ablations confirm that our results are not simply due to shrinking activations at the chosen layers.

arXiv link: <https://arxiv.org/abs/1807.06610>