Hope Speech and Help Speech: Surfacing Positivity Amidst Hate

Tackling online attacks targeting certain individuals, groups of people, or communities is a major modern-day web challenge. Research efforts in hate speech detection thus far have largely focused on identifying and subsequently filtering out negative content that specifically targets such communities. However, this blocking the hate approach alone may not suffice in certain scenarios. We focus on two important cases where amplifying the positives is equally important: refugee crisis in the era of ubiquitous internet, and heated online discussions during heightened political tension between nuclear adversaries. In the context of the Rohingya refugee crisis and the India-Pakistan conflict triggered by the Pulwama terror attack, we describe two lines of work, help speech and hope speech, exhibiting a thematic similarity of surfacing positivity amidst hate. Our work addresses several low-resource natural language processing challenges using annotation-efficient methods, Active Learning and sampling techniques, and cross-lingual sampling techniques that harness code switching.