Bringing Data Science to the Speakers of Every Language

Robert Munro
Idibon
San Francisco, CA
robert.munro@gmail.com

ABSTRACT

Speakers of more than 5,000 languages have access to internet and communication technologies. The majority of phones, tablets and computers now ship with language-enabled capabilities like speech-recognition and intelligent auto-correction, and people increasingly interact with data-intensive cloud-based language technologies like search-engines and spam-filters. For both personal and large-scale technologies, the service quality drops or disappears entirely outside of a handful of languages. Speakers of low-resource languages correlate with lower access to healthcare, education and higher vulnerability to disasters. Serving the broadest possible range of languages is crucial to ensuring equitable participation in the global information economy.

I will present examples of how natural language processing and distributed human computing are improving the lives of speakers of all the world's languages, in areas including education, disaster-response, health and access to employment. When applying natural language processing to the full diversity of the world's communications, we need to go beyond simple keyword analysis and implement complex technologies that require human-in-the-loop processing to ensure usable accuracy. In recent work where more than a million human judgments were collected on unstructured text and imagery data around natural disasters, I will present observations that debunk recent over-optimistic claims about the utility of social media following disasters. On the positive side, I will share results that show how for-profit technologies are improving people's lives by providing sustainable economic growth opportunities when they support more languages, aligning business objectives with global diversity.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*

Keywords

Natural language processing; human-computer interaction; crowdsourcing; education; disaster-response; healthcare; employment; social development

Bio

Robert is the CEO of Idibon, a company with the objective of providing language technologies for all the world's languages. In past work he has served as Chief Information Officer for the largest solar energy company in Sierra Leone; was the Chief Technology Officer for the largest use of big data technologies to track disease outbreaks globally; worked for the UN High Commission for Refugees in Liberia; lead the crowdsourced response to the 2010 earthquake Haiti; and has helped information processing in disaster response and election monitoring in more than a dozen countries. In current work, Idibon helps everyone from Fortune 500s to disaster response organizations process language data at scale. Outside of work, he has learned about the world's diversity by cycling more than 20,000 kilometers across 20 countries. Rob has a PhD from Stanford University.

KDD'14, Aug 24–27, 2014, New York, NY, USA
ACM 978-1-4503-2956-9/14/08.
http://dx.doi.org/10.1145/2623330.2630825