Location and Language Use in Social Media

Ed Chi
Google

Abstract

We now know that social interactions are critical in many knowledge and information processes. In this talk, I plan to illustrate a model-driven approach to understanding social behavior around user location and different languages in social media.

First, in 2010, we performed the first in-depth study of user location field in Twitter user profiles. We found that 34% of users did not provide real location information, frequently incorporating fake locations or sarcastic comments that can fool traditional geographic information tools. We then performed a simple machine learning experiment to determine whether we can identify a users location by only looking at contents of a user's tweets. We found that a users country and state can in fact be determined easily with decent accuracy, indicating that users implicitly reveal location information, with or without realizing it.

Second, despite the widespread adoption of Twitter in different locales, little research has investigated the differences among users of different languages. In prior research, the natural tendency has been to assume that the behaviors of English users generalize to other language users. We studied 62 million tweets collected over a four-week period. We discovered cross-language differences in adoption of features such as URLs, hashtags, mentions, replies, and retweets. We also found interesting patterns of how multi-lingual Twitter users broker information across these language boundaries. We discuss our works implications for research on large-scale social systems and design of cross-cultural communication tools.