Exploiting Contextual Information for Fine-grained Tweet Geolocation

Wen-Haw Chong, Singapore Management University
whchong.2013@phdis.smu.edu.sg

Ee-Peng Lim, Singapore Management University
eplim@smu.edu.sg

ICWSM17 – The 11th International AAAI Conference on Web and Social Media, 15-18 May 2017 @ Montreal, Canada

LARC Accepted Conference Papers: http://smu.edu.sg/centres/larc/conference-papers/
ABSTRACT

The problem of fine-grained tweet geolocation is to link tweets to their posting venues. We solve this in a learning to rank framework by ranking candidate venues given a test tweet. The problem is challenging as tweets are short and the vast majority are non-geocoded, meaning information is sparse for building models. Nonetheless, although only a small fraction of tweets are geocoded, we find that they are posted by a substantial proportion of users. Essentially, such users have location history data. Along with tweet posting time, these serve as additional contextual information for geolocation. In designing our geolocation models, we also utilize the properties of (1) spatial focus where users are more likely to visit venues near each other and (2) spatial homophily where venues near each other tend to share more similar tweet content, compared to venues further apart. Our proposed model significantly outperforms the content-only approaches.