On Predicting Religion Labels in Microblogging Networks

Minh-Thap NGUYEN, Singapore Management University
mtnguyen.2012@smu.edu.sg

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

37th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR’14), Gold Coast, Australia

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

Religious belief plays an important role in how people behave, influencing how they form preferences, interpret events around them, and develop relationships with others. Traditionally, the religion labels of user population are obtained by conducting a large scale census study. Such an approach is both high cost and time consuming. In this paper, we study the problem of predicting users’ religion labels using their microblogging data. We formulate religion label prediction as a classification task, and identify content, structure and aggregate features considering their self and social variants for representing a user. We introduce the notion of representative user to identify users who are important in the religious user community. We further define features using representative users. We show that SVM classifiers using our proposed features can accurately assign Christian and Muslim labels to a set of Twitter users with known religion labels.