

# Enrichment of User Profiles across Multiple Online Social Networks for Volunteerism Matching for Social Enterprise

Xuemeng Song  
School of Computing, National University of Singapore, Singapore  
a0095661@nus.edu.sg

## Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval

## Keywords

Volunteerism matching; Multiple sources; User profiling

## ABSTRACT

In modern society, volunteers act socially and economically as a work force of significant importance, without whom each society would face up with a major crisis. On the other hand, many talents are looking for appropriate volunteer opportunities to realize their dreams of making a contribution to the society with their expertise. This is a typical supply and demand matching issue. Fortunately, user profiling and the discovery of user volunteer tendency can benefit from users' continuous enthusiasm and active participation in diverse online social networks (OSNs) and the huge amount of publicly available user generated contents (UGCs).

Due to the different services offered by different OSNs, people tend to simultaneously participating in multiple OSNs for different purposes. Therefore, information about a user from a single source can be relatively limited. In particular, aggregating and exploring users' footprints casually left on all of these OSNs is a promising approach to generate more comprehensive summaries of users' profiles [1].

The scheme of volunteerism matching consists of two main components: *Volunteer Tendency Prediction* and *Volunteerism Matching*. Currently, we are still working on the first part. We tested our scheme on the three most popular social networks: LinkedIn, Twitter, and Facebook. The social account mapping was constructed by taking advantage of users' explicit external social links on their About.me<sup>1</sup> and Quora<sup>2</sup> profiles to their distributed profiles on multiple OSNs. In this work, we only considered users who are LinkedIn

<sup>1</sup><https://about.me/>

<sup>2</sup><http://www.quora.com/>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

SIGIR'14, July 6–11, 2014, Gold Coast, Queensland, Australia.

ACM 978-1-4503-2257-7/14/07.

<http://dx.doi.org/10.1145/2600428.2610383>.

members. In particular, we harvested the *form-based profile*, *content-based profile*, and *social connection-based profile* for each seed user from the corresponding services. We then treat the problem of volunteer tendency prediction as a binary classification task. The ground-truth for evaluation were derived from the presence of LinkedIn volunteer experience. In total, the gold dataset consists of 1,368 active volunteers and 4,005 ordinary people.

Based on the conceptual volunteering decision model proposed by Peneer et al. [2], we developed a set of features falling into four main categories: *Demographic Characteristics* (age, gender), *Personal Attributes* (user posting pattern, egocentric social networks), *Linguistic Content* (Linguistic Inquiry and Word Count features, topic distribution), and *Social Pressure & Activators* (social connections' volunteer tendency). Several state-of-the-art machine learning classification models were investigated over these features to tackle this problem, including Support Vector Machine (SVM), Gradient Boosted Decision Trees (GBDT) and Random Forest. The preliminary experimental results showed that integrating multiple OSN sources outperforms the use of only a single OSN source. Particularly, in terms of F1-Measure, performance can be improved by more than 20% through multiple OSN sources.

In summary, the findings of current work are expected to facilitate the detection of potential volunteer candidates with skills for social enterprise. Meanwhile, the finding of this work may give rise to concerns about privacy issue as expected. Although all the experiments were conducted solely based on users' publicly available data, people may still feel vulnerable being the subject of identities in our research. In the next phase of research, we will simultaneously work on privacy preserving technologies to allay users' fear of being identified and targeted; and on tackling the second component of our scheme: Volunteerism Matching. We will recommend a list of volunteer candidates for concrete demands based on users' volunteer tendency and interests. However, the establishment of ground truth for next step of research would be more challenging.

## 1. REFERENCES

- [1] M. Pennacchiotti and A.-M. Popescu. Democrats, republicans and starbucks aficionados: user classification in twitter. In *SIGKDD*, 2011.
- [2] L. A. Penner. Volunteerism and social problems: Making things better or worse? *Journal of Social Issues*, 2004.