Learner profiling beyond the MOOC platform

GUANLIANG CHEN, DELFT UNIVERSITY OF TECHNOLOGY,
DAN DAVIS, DELFT UNIVERSITY OF TECHNOLOGY,
JUN LIN, UNIVERSITY OF EINDHOVEN,
CLAUDIA HAUFF, DELFT UNIVERSITY OF TECHNOLOGY AND
GEERT-JAN HOUBEN, DELFT UNIVERSITY OF TECHNOLOGY

ABSTRACT

Large-scale learning analytics is commonly based on data traces learners generate *within* a MOOC platform such as edX during the running of a MOOC. As MOOCs typically last between five and ten weeks and many learners are rather passive consumers of the offered learning activities (Koller, 2013), this exclusive use of MOOC platform data traces severely limits the insights we can gain about our learners. This lack of data leads to coarse-grained learner profiles which in turn limit our ability to provide adaptive and personalized online learning experiences.

The social Web (where platforms such as Twitter and LinkedIn have hundreds of millions of users) potentially offers a rich source of data to supplement the MOOC platform data traces, as many learners are likely to be active on one or more social Web platforms. To investigate the benefits of *looking beyond the* MOOC *platform* for learners' data traces, we analyze the presence of 320,000 learners across 18 edX MOOCs (covering topics in computer science, business ethics, political sciences, advanced physics and renewable energies) on a number of social Web platforms including Twitter, GitHub and StackExchange. Our exploration (Chen, 2016) is guided by the following questions:

1) *On what social Web platforms can a significant fraction of MOOC learners be reliably identified?*
2) *Are learners who demonstrate specific sets of traits on the social Web drawn to certain types of MOOCs?*
3) *To what extent do social Web platforms enable us to observe (specific) user attributes that are highly relevant to the online learning experience?*

Figure 1 showcases one of our findings: we observe a significant variance in the percentage of identified learners; while on average 5% of learners could be identified in a social Web platform, in the most extreme positive case (a Functional Programming MOOC), 42% of learners could be matched to their accounts on the social coding platform GitHub.

In this work, we make the following three contributions:

- We provide a methodology to reliably and automatically identify a subset of learners from a number of social Web platforms.
- We show that it is indeed possible to derive valuable learner attributes from the social Web traces which can be used to adapt and personalize the learner experience in MOOCs.
- We show that the tracking of learners over time enables us to investigate the impact of MOOCs in the long-term.

In the talk we will not only present our approach and findings related to the three questions above but also discuss how these findings impact our understanding of massive open online learning and how we can employ them to adapt and personalize MOOC learning.

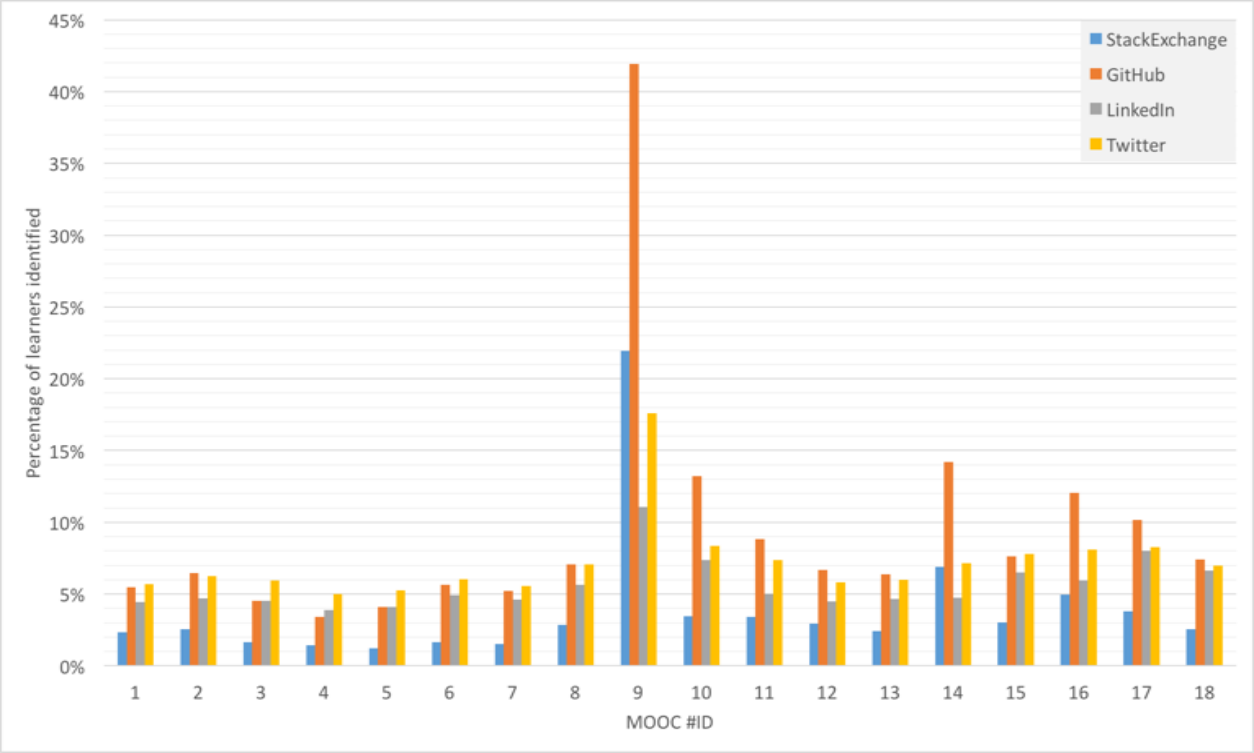

FIGURE 1: OVERVIEW OF THE PERCENTAGE OF LEARNERS IDENTIFIED ACROSS EIGHTEEN EDX MOOCS AND FOUR SOCIAL WEB PLATFORMS. THE THREE MOOCS WITH THE LARGEST FRACTION OF IDENTIFIED LEARNERS ARE MOOC #9 (*FP101X: FUNCTIONAL PROGRAMMING*), #10 (*EX101X: DATA ANALYSIS*) AND #14 (*TOPOCMX: TOPOLOGY IN CONDENSED MATTER*).

REFERENCES

Chen, G., Davis, D., Lin, J., Hauff, C., & Houben, G. J. (2016). Beyond the MOOC platform: Gaining Insights about Learners from the Social Web. In *Proceedings of the 8th International ACM Web Science Conference. To appear.*

Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: In depth. *Educause Review*, *48*(3), 62-63.