On the Impact of Personality in Massive Open Online Learning

Guanliang Chen , Dan Davis , Claudia Hauff and Geert-Jan Houben
Delft University of Technology
Delft, the Netherlands
{guanliang.chen, d.j.davis, c.hauff, g.j.p.m.houben}@tudelft.nl

ABSTRACT

Massive Open Online Courses have gained considerable momentum since their inception in 2011. They are, however, also plagued by two issues which threaten their future: learner engagement and learner retention. MOOCs regularly attract tens of thousands of learners, however, only a very small percentage of those also complete them successfully. In the traditional classroom setting, it has been established that personality impacts learning (how, what, and when to learn). An open question remains regarding to what extent this finding translates to MOOCs: do learners' personalities impact their learning (behaviour) in the MOOC setting? In this paper, we explore this question and analyse the personality profiles and learning traces of more than 700 learners taking the `Data Analysis` MOOC on the edX platform. We find personality to only weakly correlate with learning as it is captured in the MOOC setting.

Keywords

online learning, MOOC, personality

1. INTRODUCTION

Massive Open Online Courses (MOOCs) can deliver a world-class education on virtually any academic or professional development topic to any person with access to the Internet. Millions of people around the globe have signed up to courses offered on platforms such as edX[1], Coursera[2], FutureLearn[3] and Udacity[4]. At the same time though, only a small percentage of these learners (usually between 5-10%) actually complete a MOOC successfully [19], an issue that continues to plague massive open online learning. Keeping MOOC learners engaged with the course and platform are of major concerns to instructional designers and MOOC instructors alike.

Considerable research efforts have been dedicated to establish the effect of learner personality on learning in the classroom setting, e.g. [3, 21, 35, 25] & certain personality traits have been shown to be consistently correlated with learner achievement and success. Not investigated so far has been the impact of personality on learning in MOOCs — is personality predictive of success and behaviour in the current massive open online learning environments? If we were to find this to be the case, it would open avenues for personalization and adaptation of learning in MOOCs based on learners' personalities. In contrast to the classroom setting where learners form a relatively homogeneous group (in terms of age group, cultural exposure, prior knowledge), in MOOCs learners though are highly diverse [18] — a factor we hypothesize to make the subject more complex. A second question in this context is how to *estimate* the personality of learners based on MOOC data traces. The personality of learners (or users more generally) is commonly measured through self-reported questionnaires; one of the most commonly employed personality models is the so-called *Big Five personality model* [11] which is most commonly adminstered through a fifty-item self-reporting questionnaire [17]. Many learners do not take the time to fill in pre-course surveys and thus, it is also of interest to us to *estimate* learners' personality, based on their MOOC data traces alone. Such an empirical estimation of users' personality based on their digital traces has been an active area of research in the past few years, with successful predictions of personality traits based on data extracted from Facebook [16, 22, 2], Twitter [28, 15, 34, 1], Sina Weibo [14], Flickr [10] and Instagram [13]. Very diverse sets of social media traces have shown to be predictive of personality, not only behavioural (number of friends, etc.), activity and demographic features, but also image patterns and colors.

Inspired by the positive findings in these prior works, we focus in this work on the following **R**esearch **Q**uestions:

RQ1 Does personality impact learner engagement, learner behaviour and learner success in the context of MOOCs?

RQ2 Can learners' personality be predicted based on their behaviour exhibited on the MOOC platform?

We empirically investigate our research questions on 763 learners who participated in the `Data Analysis` MOOC.

We find (i) significant negative correlations between a range of behavioural MOOC features and the openness trait for

[1]`https://www.edx.org/`
[2]`https://www.coursera.org/`
[3]`https://www.futurelearn.com/`
[4]`https://www.udacity.com/`

novice learners, and (ii) significant positive correlations between behavioural features and the conscientious trait for learners with a high level of prior expertise. Overall though, we observe a lower ability to predict personality traits based user features extracted from MOOC traces, compared previous works deriving personality based on other social Web platforms.

Our empirical work also shows that the prediction of learners' personality traits based on their interactions with the MOOC platform is possible: our predictions are statistically significant for four of the five investigated personality traits and improve as more data about our learners becomes available.

2. BACKGROUND

Two strand of work come together in this work: (1) the impact of personality on learning, and (2) the prediction of personality traits based on user activities on the social Web.

2.1 Personality and Learning

Researchers in the field of Education Psychology have found each of the Big Five personality traits to be reliable predictors of academic performance (in the form of GPA) [30]. However, taking this whole body of research into account, empirical literature reviews [26] and meta-analyses [27] identify Conscientiousness as the trait with the strongest and most consistent association with academic success.

Taking individual studies into consideration, [8] found that Conscientiousness was not only a strong predictor of future academic success, it was even a more reliable predictor of a student's college GPA than his or her SAT score. And [4] also found Conscientiousness to account for more than 10% of unique variance in overall final exam marks.

Some studies, however, yield different results (no significant correlation for Conscientiousness) and find other traits to be significantly correlated with academic success. [12], for instance, found Openness and Agreeableness as the two traits most strongly correlated with academic success in a study conducted on undergraduate college students.

Other studies on education and personality, such as [6], do not concern themselves with academic success, but other factors such as a student's intrinsic motivation to attend college. They found that extroverted, agreeable, conscientious, and open students are most likely to exhibit this trait.

The above studies all employ undergraduate college students as their test subjects. However, the subjects of the present research are much more heterogeneous; given the openness of MOOCs and their accessibility, we can explore the role of personality on a new, globally diverse population of learners.

2.2 Personality Prediction based on Social Web Traces

Predicting users' personality traits based on their activities on various social Web platforms has been a very active area of research in the past years. In Table 1 we list a number of works that inspired our own investigation. The two most often considered platforms are Facebook and Twitter; they offer a myriad of diverse user traces than can be exploited for prediction purposes such as users preferences, social and academic activities, "conversations" with individuals and groups of users and so on. Especially the textual content users produce (encoded through language features) has been shown to be particularly useful to estimate users' personality, e.g. [16, 14]. Notable in Table 1 is also the diversity of the user set under investigation — ranging from a mere 71 users [1] to 180,000 users [2]. These numbers are a first pointer towards the difficulty of collecting personality ground truth data; while small user samples are gathered through questionnaires, in the two large-scale Facebook studies [22, 2] a Facebook app was developed to engage a large set of users. Studies that recruit users through crowdsourcing platforms such as Amazon Mechnical Turk, e.g. [13] may not be very reliable, due to the setup's inherent incentive for workers to quickly answer the personality questions.

Finally, Table 1 can also serve as a first indicator of the expected effectiveness of our personality predictor. The features less directly related to users (e.g. the color features in their photos or the visual patterns) yield a higher error and a lower correlation coefficient than features which are more directly related to users (the number of their friends, their use of language, etc.). Since in our scenario (personality prediction based on MOOC log traces), we also have to deal with traces which are indirectly expressing a learner's personality, we may expect our work to result in similar results as those in [10, 13].

3. PERSONALITY & MOOC DATA

Before delving into our research methodology, we briefly describe our data collection process and the specific MOOC we analyzed for this research.

3.1 MOOC

We collected personality ground truth data from learners of the `Data Analysis` MOOC — officially known as `EX101x Data Analysis: Take It to the MAX()` — which ran from August 31, 2015 to November 9, 2015 on the edX platform.

`Data Analysis` teaches various introductory data analysis skills in Excel and Python. The course was set up as an xMOOC [31]: lecture videos were distributed throughout the 10 teaching weeks. Apart from lectures, each week exercises were distributed in the form of multiple choice and numerical input questions. Each of the 146 questions was worth 1 point and could be attempted twice. Answers were due 3 weeks after the release of the respective assignment. To pass the course, $\geq 60\%$ of the questions had to be answered correctly.

Overall, 23,622 users registered for the course. Less than half of the registered learners (40%) engaged with the course, watching at least one lecture video. The completion rate was 4.75% in line with similar MOOC offerings [20].

The edX platform provides a great deal of timestamped log traces, including clicks, views, quiz attempts, and forum interactions — in the `Data Analysis` MOOC a total of 9,523,840 were recorded. We adapted the `MOOCdb`[5] toolkit to our needs and translated these low-level log traces into a data schema that is easily query-able.

3.2 Learners' Personality

We included a fifty item Big Five personality questionnaire [17] in the first week of the course as an optional component; we described our motivation for this questionnaire

[5] `http://moocdb.csail.mit.edu/`

	Data	#Users	Features	Big Five Regressor
[16]	Facebook	167	network, activities, language, preferences	$r \in [0.48, 0.65]$
[22]	Facebook	58,466	likes	$r \in [0.29, 0.43]$
[2]	Facebook	180,000	likes, status updates	RMSE $\in [0.27, 0.29]$
[28]	Twitter	335	Number of followers, following and list counts	RMSE $\in [0.69, 0.85]$
[15]	Twitter	279	language, Twitter usage, network	MAE $\in [0.12, 0.18]$
[34]	Twitter	2,927	language, Twitter usage	—
[1]	Twitter	71	Number of friends, likes, groups	MAE $\in [0.12, 0.19]$
[14]	Sina Weibo	1,766	language	$r \in [0.31, 0.40]$
[10]	Flickr	300	visual patterns	$\rho \in [0.12, 0.22]$
[13]	Instagram	113	color features	RMSE $\in [0.66, 0.95]$

Table 1: Overview of a number of past works in the area of personality prediction — shown are the platform under investigation, the number of users in the evaluation set and the type of features derived from each platform. The final column lists the evaluation metrics reported in the prediction setup: each personality trait is predicted independently, the interval shows the minimum and maximum metric reported across the five traits. The evaluation metrics are either the linear correlation coefficient (r), Spearman's rank correlation coefficient (ρ), the mean absolute error (MAE) or the root mean squared error (RMSE). The latter two are only meaningful when the normalization of the personality scores is known (here to scores between [1,5]).

in an introductory text ("aligning our education with your personality"), and did not offer any compensation.

A total of 2,195 (9.3%) registered learners began the process of filling in the personality questionnaire; 1,356 learners eventually completed this process (5.7% of registered learners). This is a common attrition rate, due to the perceived high demand (rating fifty statements) and the lack of an immediate gain for the learners.

The fifty items are short descriptive statements such as:

```
I am the life of the party.
I am always prepared.
I get stressed out easily.
```

and are answered on a Likert scale (disagree, slightly disagree, neutral, slightly agree and agree). Based on the provided answers, for each of the five personality traits (*openness*, *extraversion*, *conscientiousness*, *agreeableness*, *neuroticism*) a score between 0 and 40 is computed which indicates to what extent the learner possesses that trait.

The five traits can be summarized as follows:

- The *openness* trait is displayed by a strong intellectual curiosity and a preference for variety and novelty.
- The *extraversion* trait refers to a high degree of sociability and assertiveness.
- *Conscientiousness* is exhibited through being organized, disciplined and achievement-oriented.
- People who score high on *agreeableness* are helpful to others, cooperative and sympathetic.
- The *neuroticism* trait indicates emotional stability, the level of anxiety and impulse control.

For each learner who completed the questionnaire, we are able to compute his or her personality traits according to [17]; each learner can thus be described with a five-dimensional personality score vector.

4. APPROACH

Having gathered personality ground truth data, we now describe the features we computed for each learner based on their MOOC data traces, and the machine learning approaches employed to predict a learner's personality traits based on those features.

4.1 Features

As our work is exploratory (and to our knowledge personality prediction based on MOOC traces has not been attempted before), the features we extract are inspired by personality findings in learning outside of the MOOC setting as well as by the characteristics of the personality traits themselves.

Learners who score high on extraversion tend to have a strong need for gratification [5, 32, 23]. In the MOOC setting, such gratification can be fulfilled through interactions with other learners. The edX platform facilities interactions through their forums, and we thus explore features related to forum use. We also expect forum-based features to be useful to predict high levels of agreeableness (people who tend to help others). We hypothesize that learners who are very conscientious (i.e. have a high degree of self-organization and self-discipline) will be more disciplined in terms of video watching and quiz question answering than learners who score low in this trait, inspiring us to also explore video & quiz related features. The openness trait embodies academic curiosity and we hypothesize it to correlate positively with the amount of time spent on the platform and the material.

Concretely, we extracted the following list of twenty features for each learner (aggregating all activities throughout the running of the `Data Analysis` MOOC):

- *Time watching video material*: the total amount of time a learner as spent watching video material in the MOOCC in minutes.
- *Time solving quizzes*: the total amount of time a learner has spent on the course's quiz pages.
- *#Questions learners attempted to solve*: the total number of quiz questions the learner answered (independent of the answer being right or wrong)
- *#New forum questions*: the number of new forum posts (i.e., questions) created by the learner.

- *#Forum replies*: the number of replies (including replies to questions and comments to replies) created by the learner.
- *#Total forum postings*: the total number of new posts, replies to questions, replies to questions and comments to replies a learner created.
- *Forum browsing time*: the total amount of time the learner spent on the forum pages.
- *#Forum accesses*: the number of times the learner entered the 'Forum' page.
- *#Forum interactions*: the total number of unique learners involved in the questions the learner participated in.
- *Total time on-site*: the total amount of time (in minutes) that the learner has spent on this course's edX platform instantiation.
- *Average video response time*: the average number of minutes between a lecture video's release and the learner clicking the video's 'play' button for the first time.
- *Average quiz response time*: the average number of minutes between a quiz question's release and the learner making a first submission for it.
- *#Videos skipped*: the number of lecture videos the learner did not watch.
- *#Videos sped up*: the number of lecture videos the learner sped up during watching.
- *Maximum session time*: the maximum amount of time (in minutes) the learner spent in a single session on the course's edX site.
- *Average/standard deviation session time*: the average number of minutes/standard deviation in the learner's sessions on the course's edX site.
- *Average/standard deviation between-quizzes time*: the average number of minutes/standard deviation between answering subsequent quiz questions in the same quiz.
- *Final score*: the percentage of quiz questions the learner answered correctly at the end of the course.

Performing a correlation analysis between these features and the personality traits derived from the learners' personality questionnaires will allow us to answer **RQ1**: the extent to which personality impacts learner behaviour, engagement and success as captured through the lense of MOOC data traces.

As many of the features described here will be impacted by the learner's prior knowledge (a learner with a high amount of prior knowledge may skip many videos, while a learner without any prior knowledge may skip close to none), we distinguish two learner groups: learners with *high* prior knowledge, and learners with *low* prior knowledge. We derive a learner's level of prior knowledge based on the information provided in the general pre-course survey. In the pre-course survey, learners are asked to fill in to what degree they are familiar with certain course-specific concepts (e.g., pivot tables, named range). We aggregate learners' answers by weighting the difficulty of those concepts (given by an expert in that course) and further divide the learners into low and high prior knowledge group.

4.2 Personality Prediction

Our second goal in this work (captured in **RQ2**) is the prediction of learners' personality traits based on their MOOC data traces. To this end, we select two state-of-the-art regression models based on Gaussian Process (GP) [29] and Random Forest (RF) [24], respectively which have also been employed in previous personality prediction works, e.g. [13, 1, 14, 34].

Formally, a regression problem can be represented as $y = f(x) + \varepsilon$, where y denotes the personality trait (we predict each of the five traits independently as done in previous works), x denotes the features we derived for each learner, and ε denotes the intercept. To estimate the regression function $f(\cdot)$, GP considers the observed samples to have been drawn from a Gaussian distribution, while RF fits a number of classifying decision trees on various sub-samples and employs the averaging technique to improve the predictive accuracy. In our experiments, we set GP's noise parameter to 1.0; the number of trees in RF was set to 100.

Due to the limited number of learners, we resort to 10-fold cross-validation, performed separately for the learners with high and low prior knowledge respectively. In order to evaluate the accuracy of our personality trait predictions, we resort to Spearman's rank correlation coefficient [33] with the two variables being the learners' ground truth personality trait score (a value between 0 and 40) and the predicted trait score. Correlations are expressed as values between $[-1, 1]$ with the two boundaries indicating a perfect negative or positive alignment in ranks. Correlations close to 0 are not statistically significant and indicate that no direct relationship between the two variables exists.

5. RESULTS

In the first part of this section, we provide a basic analysis of the MOOC and the personality data we collected, and then present our findings with respect to the correlation of individual features and personality traits (Section 5.3), as well as the predictability of personality traits based on these features (Section 5.4).

5.1 Data Analysis Overview

To provide additional context of the MOOC we investigated, we provide its basic characteristics with respect to the learners that actively participated in it in Table 2. We consider a registered learner to be actively participating, if the learner clicked at least once the 'Watch' button of a lecture video. Of the more than 20,000 registered learners, this was the case for 9,493 learners — our set of "engaged" MOOC learners. Among those, about half also submitted at least one answer to a quiz question. Overall, 12% of the engaged learners earned a certificate by answering 60% or more of the quiz questions correctly. Notably, on average, less than one hour of lecture material (there is about 5 hours of lecture material in total) was consumed by the engaged learners. Less than 15% of engaged learners were active in the forum, by the end of the course, a total of 4,419 posts (questions, replies and comments) had been created.

Metrics	Results
#Learners	9,493
Completion rate	11.82%
Avg. time watching video material (in min.)	49.61
%Learners who answered at least one question	53.90%
Avg. #questions learners answered	20.89
Avg. #questions answered correctly	16.3
Avg. accuracy of learners' answers	48.25%
#Forum posts	4,419
%Learners who posted at least once	12.18%
Avg. #posts per learner	0.47

Table 2: Basic characteristics across engaged learners.

These statistics provide a first indicator of the issue we face in the prediction of personality based on MOOC log traces: data is sparse. While there are thousands of active learners, most learners are active only sporadically; only a small percentage of learners remain active throughout the entire MOOC. As already hinted at in Section 3, the MOOC we investigate is not an outlier with respect to engagement and learner success, it is rather representative of the average MOOC offered today on the major MOOC platforms.

5.2 Learners' Personalities

As stated in Section 3, we received 1,356 completed personality questionnaires from our learners. We made the design decision to present learners with the personality questionnaire at the start of the MOOC, to prevent only the most persevering subset of learners to enter our learner pool, thus decreasing bias. At the same time though, this also means that we are likely to have little activity data for most of our learners that provided us with their personality scores.

Due to the length of the personality questionnaire, we also suspect some learners to more or less randomly provide answers instead of truly *answering* to the personality statements. To investigate this effect, in Figure 1 we plot the amount of time (in minutes) it took our 1,356 learners to complete the questionnaire as extracted from the log traces. According to [17], completing this questionnaire should take between three and eight minutes, depending on a person's reading speed. We take a somewhat wider margin (Web users easily get distracted and might have been multi-tasking at the same time) and consider the personality data of all those learners as valid that spent at least three minutes and at most twelve minutes on the questionnaire. After this filtering step, we are left with 1,082 valid personality questionnaire responses that we continue to analyse in the remainder of this section.

In Figure 2 we plot the distribution of the five personality traits of those 1,082 learners. Our learners score lowest on *extraversion* and highest on *openness* and *agreeableness*. These results are in line with previous work exploring the personality of users that are active on social media [9]. The plot also shows the largest variety among our learners with respect to their *extraversion* and the smallest with respect to their *openness* to experience. These results are sensible and point to the validity of the responses — one of the defining characteristics of openness is intellectual curiosity, which every learner that starts learning through a MOOC must have to some extent. This is in contrast to the general population, where openness tends to be the trait that scores the lowest (together with extraversion), as observed for instance in [7].

We summarize the demographics of our learners with known personality traits in Table 3. The majority are male (64%) and between the ages of 20 and 40 (62%). More than 40% of our learners have completed a first university degree already.

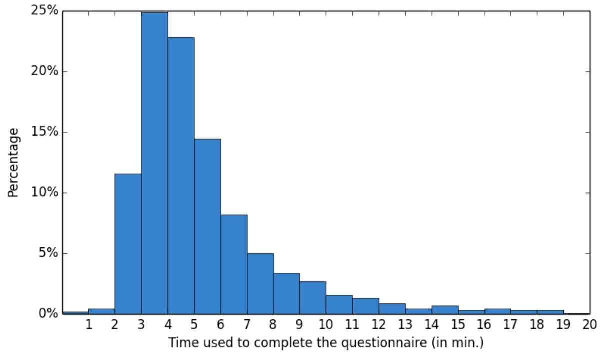


Figure 1: Overview of the fraction of learners and the time (in minutes) it took them to complete the fifty-item personality questionnaire. Only the learners that completed the whole questionnaire are included.

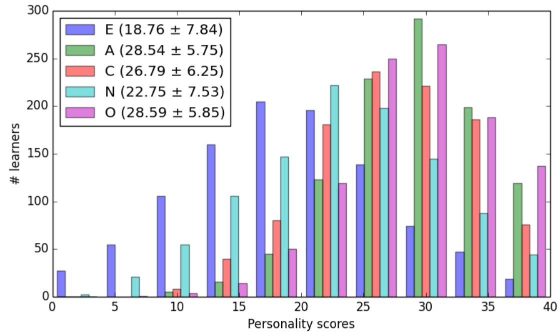


Figure 2: The histgram of the 1,082 learners' personality data. E, A, C, N, O denote Extroversion, Agreeableness, Conscientiousness, Neuroticism and Openness to experience, respectively.

Demographics	Distribution	
Gender	Female	304 (28.10%)
	Male	688 (63.59%)
	Unknown	90 (8.32%)
Age	<20	117 (10.81%)
	20 - 30	378 (34.94%)
	30 - 40	296 (27.36%)
	>40	291 (26.89%)
	Unknown	106 (9.8%)
Education level completed	Bachelor	440 (40.67%)
	Advanced degree	413 (42.75%)
	Others	133 (12.29%)
	Unknown	84 (7.76%)

Table 3: Demographic of the 1,082 learners included in our study.

5.3 Correlation Analysis

In order to conduct a meaningful correlation analyses, we partitioned our 1,082 into two sets of learners: those with high and those with low prior knowledge based on their self-reported expertise in the pre-course survey. As all questionnaires and surveys in this (and many other) MOOCs, the pre-course survey is voluntary and thus not all learners completed it. We are thus left with 763 learners who completed the personality questionnaire *and* stated their prior knowledge level.

In Tables 4 and 5 we report the measured correlation (Spearman's rank) between the features described in Section 4.1 and the learners' personality traits. As in previous works, e.g. [1, 34, 14, 2], we treat each personality trait independently. Across both sets of learners we do not observe any statistically significant correlations between behavioural features and the traits of *agreeableness* and *neuroticism*. The hypothesized increased forum activities of learners with a high agreeableness score are not supported by our data. Only two personality traits are significantly correlated with a number of features: *openness* to experience and *conscientiousness*. Among the learners with low prior knowledge (Table 4) the amount of time spent watching video lectures and number of quiz questions learners attempted are positively correlated with *conscientiousness* to a significant degree while a significant negative correlation is found for the number of videos skipped — i.e., learners with a high-self discipline and striving for achievement are likely to be more thoroughly engaged with more learning materials than learners who are not. The same features (as well as additional related features, 10 in total) are *inversely* correlated with the *openness* to experience trait to a significant degree — i.e. learners that are more intellectually curious & prefer variety are less likely to spend time focused on the learning material than learners with lower *openness* scores. As a consequence they earn a lower grade. The negative influence of this trait points to learners that are interested in a broader set of subjects (instead of steadily following a single MOOC).

In the case of learners with high levels of prior knowledge (Table 5) we observe only four significant correlations between features and personality traits: three forum features (number of replies, number of forum posts and number of forum interactions) are positively correlated with *conscientiousness*. In contrast to our expectations, learners with high levels of *extraversion* are not positively correlated with forum behaviour, in contrast, the only other significant correlation (between the amount of time spent on the forum and the *extraversion*) trait is a negative one — learners with higher levels of *extraversion* spend less time on the forum than learners with lower levels of *extraversion*.

Overall though, we have to conclude that behavioural features extracted from MOOC log traces are correlated to a lesser degree with personality than lexical or behavioural features extracted from social networks such as Facebook and Twitter, possibly due to the more constraint nature of the MOOC setting.

5.4 Personality Prediction

In this section we provide an answer to **RQ2**. We are particularly interested, to what extent we are able *early on* in the course to predict a learner's personality — if we were able to predict a learner's personality traits after one or two weeks of MOOC activities the automatic adaptation and personalization based on personality would be possible. We do not train HIGH and LOW prior knowledge learners separately, but include their prior knowledge level as an additional binary feature in the feature set.

In Figure 3 we plot for each of the personality traits the effectiveness of our two regression approaches achieve as measured by Spearman's rank correlation coefficient (as in prior works). The plots also show for each week of the course the number of active learners the personality was predicted for, with 567 active learners at the start of the course[6] (i.e. those with ground truth personality profiles) and 136 active in the last week of the course. Based on these plots, we can make a number of observations:

- significant correlations (indicating usable predictions) are achieved four of the five personality traits — the exception is agreeableness, which is not surprising, considering the correlation analysis and the lack of indicative features;
- Gaussian Processes perform better in this setting than Random Forests yielding higher correlations in three of the four traits that result in significant results;
- the correlation coefficients tend to increase with increasing course weeks as more activity data about each learner is gathered, and
- extraversion ($\rho = 0.31$) and neuroticism ($\rho = 0.22$) achieve the highest prediction accuracy by the end of the course — considering that those two traits did result in a significant correlation for only one feature in our correlation analysis, we have to conclude that more complex and higher-level features are needed to capture those traits well.

6. CONCLUSIONS

In this paper we have provided a first exploration of the relationship between massive open online learning and learners' personality traits.

Our work centered around two questions, which we evaluated in the context of the `Data Analysis` MOOC and more than 1,000 learners with valid personality profiles.

We have shown that, similar to the classroom setting, personality impacts learner engagement, behaviour and learner success (**RQ1**). We have explored a set of MOOC behavioural features and investigated their correlation with the personality traits of the Big Five personality factor model. We found various features to be correlated with the traits of *openness* and *conscientiousness* for learners with low prior knowledge. Learners with high prior knowledge exhibited fewer significant correlations, the *conscientiousness* trait was the only trait for which we observed multiple correlated features.

With respect to **RQ2** and the prediction of personality traits we can conclude that our features provide a meaningful starting point for future work — we observed significant positive correlations with all but one personality trait. The trend that over time the correlations increase (as more log traces per user become available better predictions are made) indicates the viability of the approach as well as the need to elicit more activity log traces from MOOC learners, e.g. through the offering of additional course activities and explicit guidance towards social interactions by course instructors.

[6]Note that this number is different from our 763 learners with prior expertise level and personality profile as not every user was active every week.

	E	A	C	N	O
Time watching video material (in min.)	0.00	-0.04	0.15 *	0.03	-0.18 **
Time solving quizes (in min.)	-0.02	0.02	0.07	0.02	-0.18 **
# Questions learners attempted to solve	-0.04	-0.04	0.15 *	0.03	-0.17 **
# New forum questions	0.07	0.04	0.01	0.04	0.00
# Forum replies	0.12	0.12	0.00	0.01	0.03
# Total forum postings	0.11	0.10	0.03	0.03	0.02
Forum browsing time	-0.10	0.00	0.06	-0.04	-0.13
Forum accesses	-0.11	-0.04	0.06	-0.05	-0.16 *
# Forum interactions	0.10	0.11	0.03	0.04	0.03
Total time on-site	-0.02	-0.03	0.12	0.01	-0.19 **
Average time responded to videos	0.09	-0.04	-0.04	-0.06	-0.1
Average time responded to quizs	0.03	0	-0.14	-0.1	-0.15
# Videos skipped	0.02	0.07	-0.14 *	-0.04	0.18 **
# Videos sped up	0.00	-0.03	0.10	-0.01	-0.02
Maximum session time	-0.02	-0.05	0.10	0.00	-0.17 *
Average session time	0.00	-0.03	0.05	-0.01	-0.08
Standard deviation session time	0.04	-0.01	0.11	0.02	-0.16 *
Average between-quizzes time	0.03	0.10	0.00	0.02	-0.10
Standard deviation between-quizzes time	0.01	0.06	0.04	0.01	-0.14 *
Final score	-0.06	-0.07	0.12	0.07	-0.15 *

Table 4: Overview of the Spearman's rank correlation between the 360 LOW prior knowledge learners' personality traits and their behavioural features extracted from MOOC log traces. The significant values based on two-sided t-test are marked by: * ($p < 0.01$), ** ($p < 0.001$).

	E	A	C	N	O
Time watching video material (in min.)	-0.08	-0.07	0.09	0.05	-0.01
Time solving quizes (in min.)	-0.09	-0.10	0.10	0.04	-0.03
# Questions learners attempted to solve	-0.13	-0.08	0.08	0.00	-0.03
# New forum questions	-0.04	0.04	0.10	-0.03	0.03
# Forum replies	-0.02	0.03	0.15 *	0.08	0.03
# Total forum postings	-0.03	0.02	0.15 *	0.02	0.03
Forum browsing time	-0.11	-0.04	0.02	-0.03	-0.06
Forum accesses	-0.14 *	-0.06	0.03	-0.04	-0.04
# Forum interactions	-0.02	0.02	0.15 *	0.03	0.03
Total time on-site	-0.07	-0.07	0.11	0.04	-0.03
Average time responded to videos	0.05	-0.02	-0.01	0.06	0.03
Average time responded to quizs	0.03	-0.05	0	0.05	0.02
# Videos skipped	0.09	0.08	-0.09	-0.04	0.00
# Videos sped up	0.03	0.00	0.06	0.09	0.06
Maximum session time	-0.04	-0.04	0.11	0.04	-0.04
Average session time	0.06	-0.05	0.03	0.06	-0.04
Standard deviation session time	-0.03	-0.07	0.12	0.04	-0.04
Average between-quizzes time	0.00	-0.06	0.09	0.06	0.00
Standard deviation between-quizzes time	-0.03	-0.07	0.09	0.07	0.00
Final score	-0.12	-0.05	0.07	0.01	-0.01

Table 5: Overview of the Spearman's rank correlation between the 403 HIGH prior knowledge learners' personality traits and their behavioural features extracted from MOOC log traces. The significant values based on two-sided t-test are marked by: * ($p < 0.01$), ** ($p < 0.001$).

In our future work, we will expand our analysis and exploration of behavioural features extracted from MOOC log traces for personality prediction. We will investigate human-computer interaction approaches that elicit additional log traces in MOOCs to improve the early prediction of personality traits. Most importantly, we will explore to what extent the predictions of personality allow us to automatically adapt the MOOC learning material and presentation in a meaningful manner to fulfil our ultimate goals of increasing MOOC learner engagement and success.

7. REFERENCES

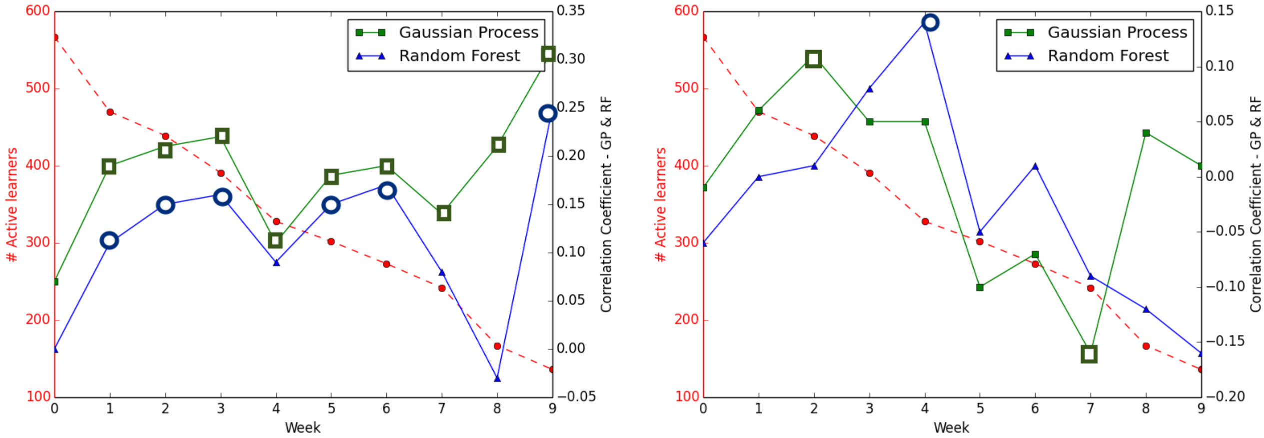
[1] S. Adali and J. Golbeck. Predicting personality with social behavior. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '12, pages 302–309, 2012.

[2] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell. Personality and patterns of facebook usage. In *Proceedings of the 4th Annual ACM Web Science Conference*, WebSci '12, pages 24–32, 2012.

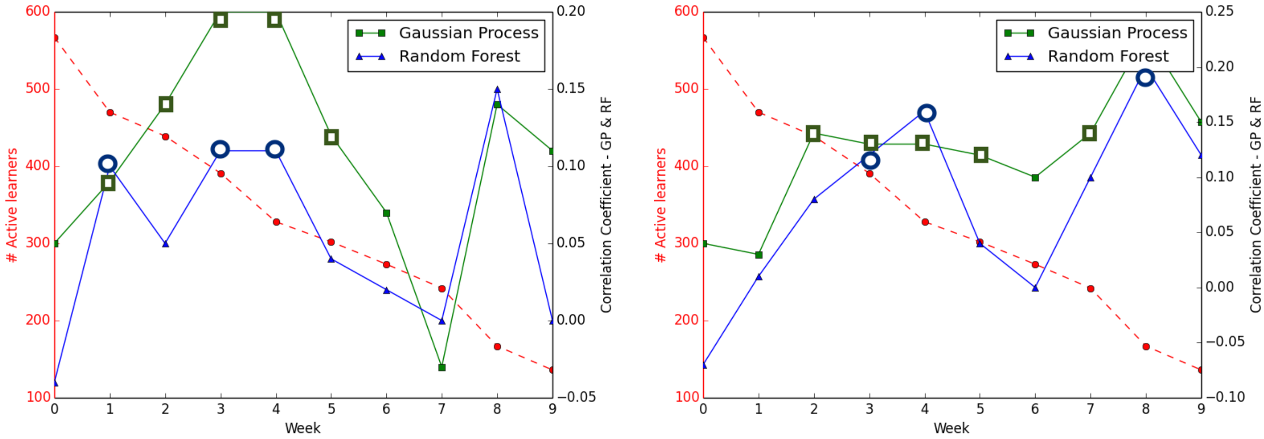
[3] G. Blickle. Personality traits, learning stratigies, and performance. *European Journal of Personality*, 10(5):337–352, 1996.

[4] T. Chamorro-Premuzic and A. Furnham. Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, 37(4):319–338, 2003.

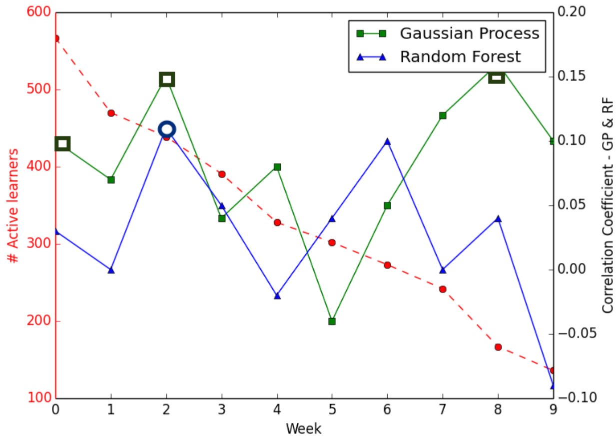
[5] S.-J. Chen and E. J. Caropreso. Influence of personality



Prediction of extraversion (left) and agreeableness (right).



Prediction of conscientiousness (left) and neuroticism (right).



Prediction of openness.

Figure 3: Overview of personality trait predictions. Each personality trait is predicted independently. In each plot, the red (dashed) line indicates the number of learners active up to course week n. The two regression-based predictors are evaluated according to Spearman's rank correlation coefficient. The empty markers denote that the corresponding results are statistically significant.

on online discussion. *Journal of Interactive Online Learning*, 3(2):1–17, 2004.

[6] M. Clark and C. A. Schroth. Examining relationships between academic motivation and personality among college students. *Learning and individual differences*, 20(1):19–24, 2010.

[7] D. A. Cobb-Clark and S. Schurer. The stability of big-five personality traits. *Economics Letters*, 115(1):11–15, 2012.

[8] M. A. Conard. Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality*, 40(3):339–346, 2006.

[9] T. Correa, A. W. Hinsley, and H. G. De Zuniga. Who interacts on the web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2):247–253, 2010.

[10] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina. Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 213–222, 2013.

[11] B. De Raad. *The Big Five Personality Factors: The psycholexical approach to personality.* Hogrefe & Huber Publishers, 2000.

[12] T. Farsides and R. Woodfield. Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual differences*, 34(7):1225–1243, 2003.

[13] B. Ferwerda, M. Schedl, and M. Tkalcic. Using Instagram Picture Features to Predict Users' Personality. In Q. Tian, N. Sebe, G.-J. Qi, B. Huet, R. Hong, and X. Liu, editors, *MultiMedia Modeling*, volume 9516 of *Lecture Notes in Computer Science*, pages 850–861. 2016.

[14] R. Gao, B. Hao, S. Bai, L. Li, A. Li, and T. Zhu. Improving user profile with personality traits predicted from social media content. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 355–358, 2013.

[15] J. Golbeck, C. Robles, M. Edmondson, and K. Turner. Predicting Personality from Twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 149–156, 2011.

[16] J. Golbeck, C. Robles, and K. Turner. Predicting personality with social media. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 253–262, 2011.

[17] L. R. Goldberg. The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26–42, 1992.

[18] P. J. Guo and K. Reinecke. Demographic differences in how students navigate through MOOCs. In *Proceedings of the first ACM conference on Learning@Scale*, pages 21–30, 2014.

[19] D. Koller, A. Ng, C. Do, and Z. Chen. Retention and intention in massive open online courses. *Educause Review*, 48(3):62–63, 2013.

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

[21] M. Komarraju, S. J. Karau, R. R. Schmeck, and A. Avdic. The big five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, 51(4):472–477, 2011.

[22] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

[23] J. Lee and Y. Lee. Personality types and learners'interaction in web-based threaded discussion. *Quarterly Review of Distance Education*, 7(1):83–94, 2006.

[24] A. Liaw and M. Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[25] S. T. McAbee and F. L. Oswald. The criterion-related validity of personality measures for predicting gpa: A meta-analytic validity competition. *Psychological Assessment*, 25(2):532, 2013.

[26] M. C. O'Connor and S. V. Paunonen. Big five personality predictors of post-secondary academic performance. *Personality and Individual differences*, 43(5):971–990, 2007.

[27] A. E. Poropat. A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, 135(2):322–338, 2009.

[28] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. Our twitter profiles, our selves: Predicting personality with twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 180–185, 2011.

[29] C. E. Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.

[30] M. Richardson, C. Abraham, and R. Bond. Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, 138(2):353–387, 2012.

[31] O. Rodriguez. The concept of openness behind c and x-moocs (massive open online courses). *Open Praxis*, 5(1):67–73, 2013.

[32] D. Rose. Personality as it relates to learning styles in online courses. In P. Resta, editor, *Proceedings of Society for Information Technology Teacher Education International Conference 2012*, pages 827–831. Association for the Advancement of Computing in Education (AACE), 2012.

[33] C. Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101, 1904.

[34] C. Sumner, A. Byers, R. Boochever, and G. Park. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 2, pages 386–393, 2012.

[35] A. Vedel. The big five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences*, 71:66–76, 2014.