Predicting PISA Scores from Students' Digital Traces

Ivan Smirnov
Institute of Education, National Research University
Higher School of Economics
Myasnitskaya ul., 20, Moscow 101000, Russia
ibsmirnov@hse.ru

Abstract

The Programme for International Student Assessment (PISA) is an influential worldwide study that tests the skills and knowledge in mathematics, reading, and science of 15-year-old students. In this paper, we show that PISA scores of individual students can be predicted from their digital traces. We use data from the nationwide Russian panel study that tracks 4,400 participants of PISA and includes information about their activity on a popular social networking site. We build a simple model that predicts PISA scores based on students' subscriptions to various public pages on the social network. The resulting model can successfully discriminate between low- and high-performing students (AUC = 0.9). We find that top-performing students are interested in pages related to science and art, while pages preferred by low-performing students typically concern humor and horoscopes. The difference in academic performance between subscribers to such public pages could be equivalent to several years of formal schooling, indicating the presence of a strong digital divide. The ability to predict academic outcomes of students from their digital traces might unlock the potential of social media data for large-scale education research.

Introduction

Measuring educational outcomes of students is crucially important for education research and policy-making. Such measurements are usually performed using standardized tests that are typically expensive and time-consuming in their development and administration. The ability to infer academic outcomes of students from their digital traces could unlock the potential of social media data for education research, and provide a way to conduct large-scale studies as became recently possible elsewhere in social science (Lazer et al. 2009).

It has already been shown that various demographic characteristics of the population such as ethnicity, gender, and income level could be inferred from tweets (Preoţiuc-Pietro et al. 2015), profile images (An and Weber 2016), user posts (Rao et al. 2011) or photographs of neighborhoods (Gebru et al. 2017). It was also shown that a wide range of personality traits including intelligence could be predicted from users' behavior on a social networking site (Kosinski, Stillwell, and Graepel 2013). As academic achievements are known to be at least as highly heritable as intelligence (Krapohl et al. 2014), one might expect that they should be predictable too.

The gold-standard instrument for evaluating educational outcomes is the Programme for International Student Assessment (PISA) (Breakspear 2014). PISA is a triennial international comparative study of student learning outcomes in reading, mathematics and science across 72 countries and economies (OECD 2016). It is arguably the most influential educational study to the point of affecting policy-making in participating countries (Egelund 2008; Ertl 2006; Rautalin and Alasuutari 2009). In this paper, we predict PISA scores of individual students from information about their activity on a popular social networking site.

We use data from the Russian panel study "Trajectories in Education and Career" that tracks 4,400 students who participated in PISA in 2012. In addition to survey data, participants' online activity information was collected in 2016. We assume that academic performance is a relatively stable characteristic and that the time interval of four years between two measurements should not prevent our ability to make a prediction.

The information about the online behavior of students was collected from VK (a Russian analogue of Facebook). VK is ubiquitous among young Russians: more than 90% of 18–24 years old use it regularly (Public Opinion Foundation 2016). Users of this social network are subscribed to various public pages that might be dedicated to anything from a local bar to a theater and from handcraft to quantum physics. There are more than 28 million public pages on VK. In our sample, the total number of different public pages is 73,000, and the median number of users' subscriptions is 54. We use a dimensionality reduction technique to extract ten components that represent interests of each user. We then build a linear regression model to predict their PISA scores.

Results

Predicting PISA scores

PISA scores are scaled so that the OECD average in each domain (mathematics, reading, and science) is 500 and the standard deviation is 100, while 40 score points roughly correspond to the equivalent of one year of formal schooling (OECD 2014a). Instead of a single estimate of students'

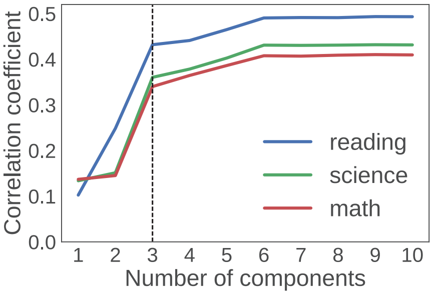

Figure 1: Pearson correlation coefficient between predicted and real PISA scores as a function of the number of components used in the linear regression. The results are shown for three PISA subjects (reading, mathematics, and science). The 10-fold cross-validation was used to control for overfitting. The largest increase in model performance is provided by the third (*academic*) component of users' interests. The second component is correlated with users' gender and provides a substantial increase in model performance for the reading score.

abilities, PISA provides five plausible values drawn from an estimated distribution of their outcome level (see (OECD 2014b) for details about PISA's plausible values and Methods for their use in our analysis).

We use singular value decomposition to extract ten main components containing information about users' subscriptions to public pages (see Methods for details). We then use a linear regression model to predict PISA scores and compute the Pearson correlation coefficient between predicted and real outcomes as a measure of model performance. The 10-fold cross-validation is used to control for potential overfitting of the model. The average correlation coefficient as a function of a number of components used in the regression is shown in Fig. 1.

The largest increase in performance is provided by the third component that is correlated with PISA scores for all three subjects (r = 0.35 for reading, r = 0.33 for science, and r = 0.30 for mathematics). The third component, therefore, might be considered as an *academic component* of users' interests.

The substantial increase in model performance for reading scores is also provided by the second component. The names of public pages that contribute most to this component (i.e. the pages with highest absolute values of weights) suggest that it might reflect users' gender (see Table 1). Indeed, the second component is strongly correlated with gender (r = 0.66). It is no surprise that *gender component* provides a substantial increase in model performance for reading scores, because there is a large gender gap in reading: girls outperform boys in this subject in every country and economy where PISA is administered (OECD 2014a).

Digital divide

With almost universal Internet penetration in developed countries, concerns about the digital divide have shifted from the simple question of access to the broader notion of inequalities in the usage of the Internet (DiMaggio and Hargittai 2001). It was observed that higher-educated people more often use the Internet for educational purposes, while less-educated people more often use it for entertainment (Pearce and Rice 2017; Büchi, Just, and Latzer 2016; Van Deursen and Van Dijk 2014). We find a similar pattern in subscriptions to VK public pages.

Table 1: Names of public pages that contribute most to *gender component* of users' interests. If the name of a public page is not self-descriptive, then the main topic of its content is described in parentheses. Translated from Russian.

Negative contribution	Positive contribution
40 KG	Orlyonok (funny pictures, obscene language)
90-60-90 Sport girls	MDK (funny pictures, obscene language)
Charm School	IGM (computer games)
Beauty School	Academy of Decent Guys
Modern Girl	AUTO

We select public pages that contribute most to the identified *academic component* and compute average PISA scores for its participants (Table 2). Names of the pages with the positive contribution to the component are related to science and art, while pages with highest negative weights concern humor and horoscopes.

These pages were not selected to provide the highest difference in PISA scores, instead they arise exclusively from the structure of users' interests. Despite this fact, the observed gap in performance of subscribers to these public pages could be dramatic. The subscribers to the *World Arts and Culture (WAC)* page demonstrate results on par with top-performing countries and outperform subscribers to *Love Horoscope* by 79–88 score points that are roughly equivalent to two years of formal schooling. Note that these are not marginal public pages on VK, with almost two million subscribers to the *WAC* page and more than four million subscribers to *Love Horoscope*.

Predicting proficiency levels

To help users interpret what student scores mean in substantive terms, the PISA scale is divided into six proficiency levels (OECD 2014a). In Table 3, we report the ability of our model to distinguish between students of different proficiency levels in reading. The performance is measured as the area under the ROC curve (AUC). The results for reading are slightly better than results for science and mathematics.

According to the OECD, Level 2 is a baseline proficiency that is required to participate fully in modern society (OECD 2014a). Students who do not meet this baseline are considered as low-performing. High-performing students are those who achieve proficiency Level 5 or higher. If this proficiency is achieved in all three subjects simultaneously, then students are able to draw on and use information from multiple and indirect sources to solve complex problems, and they

Table 2: Names of public pages that contribute most to the *academic component* of users' interests. If the name of a public page is not self-descriptive, then the main topic of its content is described in parentheses. Names are translated from Russian. Mean values of subscribers' scores with standard errors (in parentheses) are provided for each of three PISA subjects.

	Math	Reading	Science
Positive contribution			
WAC (World Arts and Culture)	538 (4.6)	530 (4.5)	532 (4.3)
Science	521 (4.2)	502 (4.1)	516 (3.8)
Best poems of great poets	509 (4.0)	507 (4.0)	508 (3.9)
Science and Technology	507 (4.1)	479 (4.3)	504 (4.0)
Five Best Movies	505 (3.9)	492 (3.9)	503 (3.7)
Negative contribution			
F*CK (funny pictures often related to sex)	473 (3.3)	449 (3.4)	472 (3.2)
Killing humor	471 (5.1)	447 (5.1)	471 (4.7)
Cool Gags	467 (4.9)	444 (5.1)	465 (4.9)
Unorthodox Horoscope	462 (5.1)	450 (5.3)	460 (5.0)
Love Horoscope	450 (5.3)	442 (5.8)	453 (5.2)

will be at the forefront of a competitive, knowledge-based global economy (OECD 2014a).

The ability for our model to distinguish between low- and high-performing students is 0.90 for mathematics, 0.92 for science, and 0.94 for reading.

Discussion

We show that there is a relatively strong signal with respect to academic performance in data on students' subscriptions to various public pages on the VK social networking site. The identified *academic component* of users' interest explains as much variation in their learning outcomes as the socioeconomic status that is measured by the PISA index of economic, social and cultural status (ESCS) (OECD 2013). While the obtained model does not allow reconstruction of raw scores reliably, it is able to discriminate between low- and high-performing students with a high degree of accuracy. Unlike previous research that focused on predicting academic achievements of students from one university (Kassarnig et al. 2017; Lian et al. 2016; Wang et al. 2015), we were able to use data on a representative sample of Russian students.

It should be taken into account that PISA is not an optimal instrument for the measurement of individual performance because its plausible values contain random error variance components. Instead, PISA was designed to provide accurate summary statistics about the population of interest within countries and about correlations between key variables (Jerrim et al. 2017). That fact might affect the performance of our model.

We also find a significant digital divide among students. Subscribers to the *World Arts and Culture (WAC)* page demonstrate results that are much higher than results of their peers subscribed to *Love Horoscope*. The gap is roughly equivalent to two years of formal schooling. The large variety of public page names (from *Cool Gags* to *Science and Technology*) masks the fact that their content is rather homogenous. Most of the pages, including *WAC*, post predominantly funny pictures and videos. Thus, there is little reason to believe that subscription to these pages might significantly affect the academic performance of students. However, the subscription to pages, the names of which suggest a relation to science and art, might serve a signaling purpose and play a role in students' self-identification. Here, *WAC* provides an illustrative example because its tagline reads "Only intellectuals. Only hardcore".

The rate of adoption of large-scale datasets enabled by social media platforms is slower in education research than in some other areas of social science. This slow rate of adoption might be partly explained by the fact that one of the main variables of interest (i.e. academic performance) is generally not available in these datasets to the researchers. The ability to infer the academic performance of social media users, even if imperfectly, might open new possibilities for educational research. It also raises ethical concerns about the potential use of such data. It has already been reported that some universities used VK data to target potential entrants by identifying high school graduates with "a keen interest to humanities" and planned to include prediction of "intelligence, creativity, and motivation" in the upcoming year (Tomsk State University 2017). While screening candidates in the hiring process has already been discussed in the literature (Landers and Schmidt 2016; Drouin et al. 2015; McDonald, Thompson, and O'Connor 2016), these new practices require additional attention.

Methods

Sample

We use data from the Russian Longitudinal Panel Study of Educational and Occupational Trajectories (TrEC) (Kurakin 2014). The study tracks 4,399 students from 42 Russian regions who took the PISA test in 2012. In 2016, publicly available information from the social networking site VK was collected for 3,483 TrEC participants who provided informed consent for the use of this data for research purposes. Note that while the initial sample was representative of 9th-grade high school Russian students in 2012, the social network data is not necessarily representative.

Table 3: Model performance in discrimination between different proficiency levels in reading measured as the area under the ROC curve.

	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Bellow level 1	0.598	0.715	0.840	0.922	0.954	0.992
Level 1		0.631	0.775	0.880	0.926	0.980
Level 2			0.653	0.785	0.845	0.931
Level 3				0.655	0.730	0.861
Level 4					0.579	0.739
Level 5						0.674

VK Data

VK is the largest European social networking site, with more than 100 million active users. It provides an application programming interface (API) that allows systematic downloading of public information from users' profiles, including their subscriptions to various public pages. The users from our sample are subscribed to 73,389 different public pages. The median number of users' subscription is 54. The maximum value is 1,000, because this is the largest number of pages that could be returned by the VK API. We exclude pages that have less than ten subscribers and users with less than ten subscriptions, resulting in a sample of 2,637 users and 4,485 pages.

PISA

Instead of a single point estimate of students' abilities, PISA provides five "plausible values" that are described in detail in its technical report (OECD 2014b). We deal with these values in the following way. First, we build five linear regression models for each of the plausible values and average their coefficients to obtain the final model. We then measure the performance of the resulting model for each of the five plausible values and report its average performance.

Model

From information about users' subscriptions to public pages, we construct matrix A 2,637 x 4,486 so that $A_{ij} = 1$ if user i is subscribed to public page j, and $A_{ij} = 0$ otherwise. We then use singular value decomposition of A to extract ten main components representing user's interests. These components were then used in a linear regression model to predict users' PISA scores. The 10-fold cross-validation was used to control for model overfitting. This approach is similar to the one used for Facebook data (Kosinski, Stillwell, and Graepel 2013) and makes the comparison of results possible. The use of non-linear models or increasing the number of components do not provide any substantial increase in the model's performance.

Acknowledgments

The data of the Russian panel study "Trajectories in Education and Career" (TrEC http://trec.hse.ru/) are presented in this work. Support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged.

References

An, J., and Weber, I. 2016. # greysanatomy vs.# yankees: Demographics and hashtag use on twitter. In *Tenth International AAAI Conference on Web and Social Media*.

Breakspear, S. 2014. How does PISA shape education policy making? Why how we measure learning determines what counts in education. In *Centre for Strategic Education Seminar Series Paper*, volume 40.

Büchi, M.; Just, N.; and Latzer, M. 2016. Modeling the second-level digital divide: A five-country study of social differences in internet use. *New media & Society* 18(11):2703–2722.

DiMaggio, P., and Hargittai, E. 2001. From the digital divideto digital inequality: Studying internet use as penetration increases. *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University* 4(1):4–2.

Drouin, M.; OConnor, K. W.; Schmidt, G. B.; and Miller, D. A. 2015. Facebook fired: Legal perspectives and young adults opinions on the use of social media in hiring and firing decisions. *Computers in Human Behavior* 46:123–128.

Egelund, N. 2008. The value of international comparative studies of achievement–a danish perspective. *Assessment in Education: Principles, Policy & Practice* 15(3):245–251.

Ertl, H. 2006. Educational standards and the changing discourse on education: The reception and consequences of the pisa study in germany. *Oxford Review of Education* 32(5):619–634.

Gebru, T.; Krause, J.; Wang, Y.; Chen, D.; Deng, J.; Aiden, E. L.; and Fei-Fei, L. 2017. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences* 114(50):13108–13113.

Jerrim, J.; Lopez-Agudo, L. A.; Marcenaro-Gutierrez, O. D.; Shure, N.; et al. 2017. What happens when econometrics and psychometrics collide? An example using PISA data. *Department of quantitative social science* 17–04.

Kassarnig, V.; Bjerre-Nielsen, A.; Mones, E.; Lehmann, S.; and Lassen, D. D. 2017. Class attendance, peer similarity, and academic performance in a large field study. *PloS one* 12(11):e0187078.

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

Krapohl, E.; Rimfeld, K.; Shakeshaft, N. G.; Trzaskowski, M.; McMillan, A.; Pingault, J.-B.; Asbury, K.; Harlaar, N.; Kovas, Y.; Dale, P. S.; et al. 2014. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences* 111(42):15273–15278.

Kurakin, D. 2014. Russian longitudinal panel study of educational and occupational trajectories: Building culturally-sensitive research framework.

Landers, R. N., and Schmidt, G. B. 2016. Social media in employee selection and recruitment: An overview. In *Social Media in Employee Selection and Recruitment*. Springer. 3–11.

Lazer, D.; Pentland, A. S.; Adamic, L.; Aral, S.; Barabasi, A. L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; et al. 2009. Life in the network: the coming age of computational social science. *Science* 323(5915):721.

Lian, D.; Ye, Y.; Zhu, W.; Liu, Q.; Xie, X.; and Xiong, H. 2016. Mutual reinforcement of academic performance prediction and library book recommendation. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, 1023–1028. IEEE.

McDonald, P.; Thompson, P.; and O'Connor, P. 2016. Profiling employees online: shifting public–private boundaries in organisational life. *Human Resource Management Journal* 26(4):541–556.

OECD. 2013. *PISA 2012 Results (Volume II): Excellence through Equity Giving Every Student the Chance to Succeed*. OECD Publishing.

OECD. 2014a. *PISA 2012 Results: What Students Know and Can Do Student Performance in Mathematics, Reading and Science*. OECD Publishing.

OECD. 2014b. *PISA 2012 Technical report*. OECD Publishing.

OECD. 2016. *PISA 2015 Results (Volume I): Excellence and Equity in Education*. OECD Publishing.

Pearce, K. E., and Rice, R. E. 2017. Somewhat separate and unequal: digital divides, social networking sites, and capital-enhancing activities. *Social Media and Society* 3(2):2056305117716272.

Preoţiuc-Pietro, D.; Volkova, S.; Lampos, V.; Bachrach, Y.; and Aletras, N. 2015. Studying user income through language, behaviour and affect in social media. *PloS one* 10(9):e0138717.

Public Opinion Foundation. 2016. Online practices of russians: social networks. http://fom.ru/SMI-i-internet/12495. [Accessed 06.01.2018].

Rao, D.; Paul, M.; Fink, C.; Yarowsky, D.; Oates, T.; and Coppersmith, G. 2011. Hierarchical bayesian models for latent attribute detection in social networks. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*.

Rautalin, M., and Alasuutari, P. 2009. The uses of the national pisa results by finnish officials in central government. *Journal of Education Policy* 24(5):539–556.

Tomsk State University. 2017. The accuracy of TSU program in finding "matching" entrants in social networks is 82%. http://goo.gl/7PA8EP. [Accessed 11.01.2018].

Van Deursen, A. J., and Van Dijk, J. A. 2014. The digital divide shifts to differences in usage. *New media & Society* 16(3):507–526.

Wang, R.; Harari, G.; Hao, P.; Zhou, X.; and Campbell, A. T. 2015. Smartgpa: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 295–306. ACM.