

Ubiquity Symposium

MOOCs and Technology to Advance Learning and Learning Research

Offering Verified Credentials in Massive Open Online Courses

by Andrew Maas, Chris Heather, Chuong (Tom) Do, Relly Brandman, Daphne Koller, and Andrew Ng

Editor's Introduction

Massive open online courses (MOOCs) enable the delivery of high-quality educational experiences to large groups of students. Coursera, one of the largest MOOC providers, developed a program to provide students with verified credentials as a record of their MOOC performance. Such credentials help students convey achievements in MOOCs to future employers and academic programs. This article outlines the process and biometrics Coursera uses to establish and verify student identity during a course. We additionally present data that suggest verified certificate programs help increase student success rates in courses.

Ubiquity Symposium

MOOCs and Technology to Advance Learning and Learning Research

Offering Verified Credentials in Massive Open Online Courses

by Andrew Maas, Chris Heather, Chuong (Tom) Do, Relly Brandman, Daphne Koller, and Andrew Ng

Coursera has partnered with more than 100 top universities and institutions, among them Princeton, Caltech, Johns Hopkins, University of Pennsylvania, and others, to offer free courses online for anyone. Since announcing our initial partnerships in April 2012, Coursera has enrolled more than 5 million students across every country in the world.

To help provide students a tangible benefit for their achievement, every student who successfully completes a Coursera course can earn a Statement of Accomplishment (SoA), a letter from the instructor to the student. However, a common concern regarding any form of credentialing in an online environment is that of student identity: How do we know the student whose name appears in the credential is the one doing the work in the course? To help address this concern, Coursera has recently established the “Signature Track,” a process by which students link their coursework, done completely online, to their real identity. After successfully completing a course in the Signature Track, students are given a Verified Certificate issued to them under their real name, and authorized by both the participating university and Coursera. The certificate has a unique verification code, allowing the student to verify to a third-party (such as an employer) their successful completion of the course. Certificates are a stronger credential than alternatives like Mozilla badges because the certificate is tied to real-world identity. Universities, through Coursera, trust that the named student completed all coursework.

Indeed, motivations for enrolling in Signature Track often relate to credentialing; many Signature Track students are interested in the opportunity to share a trusted statement of their class performance with educational institutions and employers. Moreover, as we discuss at the end of this article, participation in the Signature Track is associated with another benefit: A significantly increased retention rate for students in the course, even after correcting for the students' initial level of motivation.

Participation in the Signature Track costs a modest fee (currently around \$30-70). This participation is entirely optional; access to the content remains free, and students who successfully complete the class continue still receive the usual SoA. For those students who cannot afford the Signature Track fee, we have put in place a Financial Aid program. This program requires students to submit a financial aid application explaining why they would benefit from the certificate but cannot afford it. Students approved for the financial aid program take the course in signature track at no cost. As of June 2013, more than 2,500 students have participated in the Financial Aid program.

Identity Verification and Authentication

Establishing students' identity and participation in their educational experience is a challenging proposition in online courses where the instructors and students do not meet face to face. Coursera faces the additional challenge of scale: A typical Coursera course has 40,000 - 60,000 students, and thus the processes set up to support Signature Track need to work efficiently with little or no intervention from staff or instructors. We have identified two facets to establishing identity in online education:

1. **Identity verification.** Matching the student's registered identity to a real-world identity, and
2. **Identity authentication.** Ensuring the person who completes the weekly coursework is the same student who originally enrolled.

Note the verification and authentication problem here is more challenging than what is required in many web-scale platforms. In online banking and shopping a user never benefits when another person successfully uses his or her identity. Students, however, may want others to complete their work for them, and thus may collude with intruders to defeat the identity

system. Coursera uses two biometric authentication approaches, face photos and typing patterns, to establish and maintain the link between a student account and real-world identity. Initially identity verification occurs to establish a connection between the student and his or her Coursera account. Coursera performs identity verification during the Signature Track enrollment process early in the course. Enrollment requires the student to provide a webcam photo of his or her face as well as a photo of a government-issued ID document, such as a passport or driver's license. A Coursera employee verifies the picture and name on the ID photo match the student's face photo and name. Following this verification, the ID photo is deleted for privacy reasons. Enrollment additionally requires that the student establish a "keystroke biometric profile" by typing a short phrase. Throughout the course, the Coursera platform performs identity authentication, which verifies that the person completing course exercises is indeed the student who enrolled in Signature Track.

The original face photo and typing sample serve as reference points to authenticate the student as he or she completes work in courses that offer Signature Track. When completing assignments in these courses, the student provides a typing sample and face photo to maintain the connection with the identity associated with his or her account. Coursera currently offers 53 Signature Track courses with multiple exercises and assignments in each course. Coupled with the growing number of students enrolled in Signature Track, it is infeasible for Coursera employees to personally verify each assignment by each student. Instead, we built an automated typing pattern biometric verification system and tools to prioritize Coursera employee time in verifying face photo matches.

Identity authentication with keystroke dynamics. Face photo matching by human annotators is generally a reliable method for identity authentication. For example, Kumar et al. find human annotators achieve 99 percent recognition accuracy on a challenging dataset of 60,000 face images [1]. However, students' access to a webcam varies across devices, and lighting conditions are not always suitable for capturing a face photo of sufficient quality. We therefore built a "keystroke dynamics" identity authentication system to reliably perform authentication requiring only a keyboard, without the need for a human annotator. Keystroke dynamics relies on the unique rhythms and cadence of keypress events, which occur as students type a given phrase. Keystroke biometric systems are widely used to reinforce computer security in banking and enterprise settings [2].

We designed the Coursera keystroke authentication system to be minimally intrusive to students' learning experience while still providing accurate identity authentication. To obtain a typing sample, the student is asked to transcribe a short phrase. At enrollment time, and often during identity authentication, the student transcribes: "I certify this submission as my own original work completed in accordance with the Coursera Honor Code." A Javascript tool based on the open-source library from [Keytrac](#) records the time in milliseconds of each key-down and key-up event relative to the beginning of the typing session. The time resolution of this system is thus limited by the capabilities of the student's web browser and operating system. While this resolution is lower than that of a system that logs events within the operating system, we have found it sufficient in practice and does not require students to install additional software.

The raw stream of keypress data is not suitable for direct use with our machine learning classification system. Keypress events in isolation are highly variable and are affected by random environmental factors, such as interruptions. Furthermore, the sequence of keys pressed varies due to typos and their corrections, as well as key choices such as using "caps lock" versus "shift" to capitalize letters. We transform the raw typing data into a set of features, which accurately represents the keystroke dynamics while abstracting away less important information such as time delays due to interruption. Building on previous work [3], we represent a typing stream by measuring time differences among small groups of key presses. Time differences between subsequent keys are strong indicators of an individual's unique typing pattern, and generalize sufficiently across different typing sessions.

Figure 1 shows a scatter plot of 20,000 typing samples from 3,122 students embedded in two dimensions by the t-SNE algorithm [4]. Similarities among typing features correspond roughly to distance in the 2-D space. We selected two students and colored all sample from each student differently from other points. Typing samples from the same student tend to cluster together, indicating that students produce repeatable typing patterns. Further, we notice some macroscopic structure of the embedded points, which corresponds to broad similarities among groups of students.

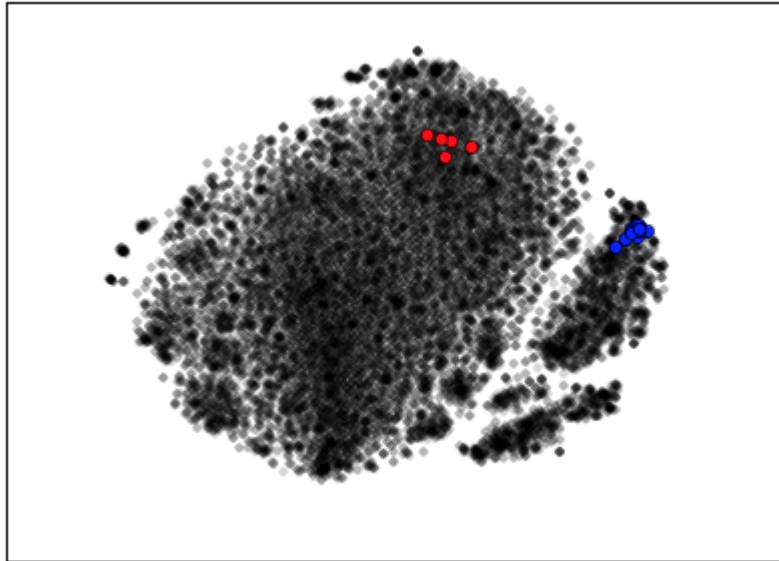


Figure 1. 20,000 typing samples from 3,122 students projected into two dimensions by the t-SNE algorithm. All typing samples from two students are colored differently (blue and red) to illustrate the grouping of typing samples of individuals relative to the overall population.

Because of the large student population in Signature Track, we can accurately train a system to obtain precise authentication performance. We train our system on millions of valid and invalid typing sample pairs. After setting an appropriate decision threshold, the trained system correctly recognizes valid authentication attempts while allowing few imposter authentication attempts to pass undetected.

At enrollment time, a student types a given phrase twice and we run an authentication attempt to ensure our system does indeed recognize the two samples as coming from the same student. When completing assignments, the student is prompted to transcribe a phrase. Our system then checks whether this new typing sample matches against the typing samples recorded during enrollment. The system additionally implements a number of heuristic checks to prevent abuses such as attempts to intentionally mimic the typing patterns of other users through manual or automated means.

Identity authentication with face photos. Authentication attempts using solely keystroke dynamics are not easily interpretable by a human. To facilitate an interpretable secondary authentication mechanism, our system takes a face photo via the webcam of the student as she

or he completes an assignment. These photos allow an annotator to easily verify identity in cases where the keystroke-based system is uncertain. As Coursera continues to grow, it becomes more challenging for human annotators to verify all face photos for a student whose verified status is in question. To facilitate the process, we are building automated face matching tools to prioritize annotator time to check only the most uncertain cases. Accuracy of automated face verification systems is steadily improving and can already save annotator time by filtering clear matches [1].

Overall authentication requirements. For a student taking a verified course, we require a minimum percentage of the student's completed assignments have a successful identity authentication check. A successful check can come from either a match of typing pattern as determined by the keystroke dynamics system or from a face photo match evaluated by a human annotator. At the conclusion of the course, this authentication data is reviewed, along with the student's score. If the student has both met the course passing criteria and authenticated their coursework, they are issued a Verified Certificate from the participating university and Coursera.

It is important to clarify the scope of the assurance that Signature Track provides. As with any authentication mechanism, the system described here does not provide complete protection against individuals who are sufficiently determined to defeat the online identity verification mechanisms employed. Furthermore, the identity verification provided by Signature Track is not a full-fledged approach for ensuring academic integrity. In particular, it is possible for a student to find answers to questions on the web or to get help from a friend. Nevertheless, this level of assurance is comparable to that of any work assigned to students at home, even in an on-campus class. It is analogous to a handwriting verification (or "signature") for a student's work done at home. In high-stakes settings where a more rigorous assurance of academic integrity is necessary (such as classes that confer college credit), we employ an additional mechanism for assessment that uses remote [digital proctoring](#) via webcam, microphone, and a locked-down workstation.

Signature Track and Retention

Signature Track is a relatively new process for MOOCs. It is important to understand whether the process affects student learning outcomes beyond the effects of earning a verified certificate. In the first nine classes to offer the Signature Track program, we conducted a survey soliciting feedback regarding student intentions and likelihood of completing the course. The survey, which was administered after the close of registration for the Signature Track program in each class, had a 31 percent response rate among Signature Track students and 6 percent response rate among non-Signature Track students. Among all students in the nine classes studied, Signature Track participants were far more likely to complete their respective classes than non-Signature Track students (63 percent versus 3 percent; see Figure 2). This finding is not surprising: naturally, the population of students who sign up for Signature Track is skewed towards individuals who are more motivated and committed to their courses than the average student. Therefore, one might ask whether the differences in completion rates are simply due to this selection bias rather than the actual influence of Signature Track participation. In other words, does participation in Signature Track actually improve completion rates?

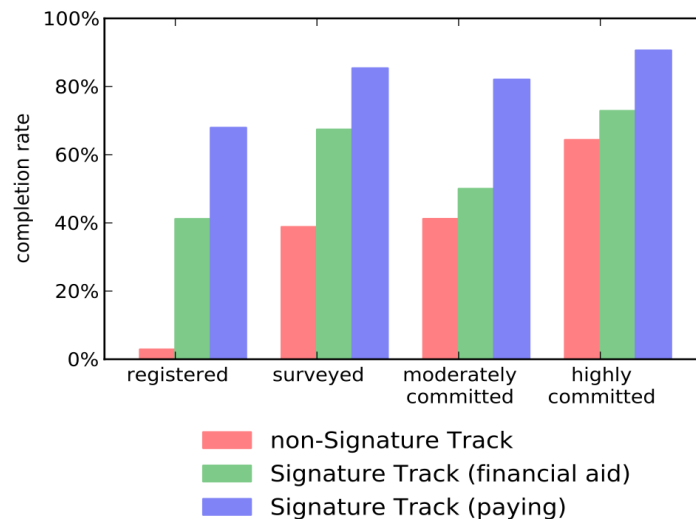


Figure 2. Completion rates in the Signature Track program. The groups of bars correspond to all registered students in each class, all survey respondents, students who indicated a moderate level of commitment to completing the course, and students who indicated a high level of commitment to completion. Within each group, students are distinguished according to whether they did not sign up for the Signature Track program, whether they signed up for financial aid in the Signature Track program, or whether they signed up as a paying Signature Track student.

Additional study suggests the answer is yes: When restricting to only the most highly committed students (as assessed through the survey), Signature Track participants show a marked improvement in completion rates compared to non-Signature Track participants (88 percent versus 64 percent). A similar difference holds among moderately committed students (78 percent versus 41 percent), suggesting that saturation of the “commitment” scale cannot be the explanation for this improvement. One possible explanation is the importance of making a financial commitment as an element of the Signature Track program. Indeed, we find that overall completion rates also differ between financial aid and paying students (41 percent versus 68 percent). However, this last finding must be interpreted with care, since differences in demographics, level of preparedness upon entering the course, general life circumstances, and unrelated extracurricular responsibilities are potential confounding factors.

Conclusions and Implications

Over 70 percent of individuals taking Coursera courses have a bachelor’s degree or higher. Many of these students use these courses as a way to enhance their career. For example, a VP of Engineering at an international mobile communications company took [Georgia Tech’s “Computational Investing”](#) course in preparation for a possible career transition into quantitative trading. A recently graduated data scientist took the same course to help with his investing startup. An individual employed in the Ministry of Economy of the Kyrgyz Republic is taking [Penn’s “An Introduction to Operations Management”](#) to help with a project aimed at improving foreign trade procedures for customs. Finally, another individual, currently working as a tutor, took [“Genetics and Evolution” through Duke](#) in preparation for becoming a high school science teacher.

All of these students used the Signature Track as a way of demonstrating their achievement to their current or prospective employers. Currently, only a small fraction of one’s adult working life is spent advancing one’s formal education. Our aim with Signature Track is to open formal education opportunities to more aspects of everyday life. We imagine a world where a student can add a custom blend of top-quality courses to their core education credentials (such as a bachelor’s degree) to continue to learn and grow throughout their lives.

References

- [1] Kumar, N., Berg, A. C., Belhumeur, P. N., and Nayar, S. K. (2009, September). Attribute and simile classifiers for face verification. In *IEEE 12th International Conference on Computer Vision, 2009*. IEEE, New York, 2009, 365-372.
- [2] Moskovitch, R., Feher, C., Messerman, A., Kirschnick, N., Mustafic, T., Camtepe, A., Bernhard Löhlein, B., Ulrich, H., Möller, B., Lior, R., and Elovici, Y. Identity theft, computers and behavioral biometrics. In *ISI'09. IEEE International Conference on Intelligence and Security Informatics, 2009*. IEEE, New York, 2009, 155-160.
- [3] Gunetti, D. and Picardi, C. Keystroke analysis of free text. *ACM Transactions on Information and System Security (TISSEC)* 8, 3 (2005), 312-347.
- [4] Van der Maaten, L. and Hinton, G. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 85 . (2008), 2579-2605.

About the Authors

Andrew Maas is a Ph.D. candidate in computer science at Stanford University and a software engineer at Coursera. He created Coursera's keystroke biometric.

Chris Heather has a BS in business from the University of Virginia and does business development at Coursera.

Chuong (Tom) Do has a Ph.D. in computer science from Stanford University and is a software engineer at Coursera.

Relly Brandman has a Ph.D. in chemical and systems biology from Stanford University and does course operations at Coursera.

Daphne Koller is co-CEO and co-Founder of Coursera. She is the Rajeev Motwani Professor in the Computer Science Department at Stanford University.



Andrew Ng is co-CEO and co-Founder of Coursera. He is the Director of the Stanford Artificial Intelligence Lab and Associate Professor of Computer Science Department at Stanford University.

DOI: 10.1145/2591684