International Review of Research in Open and Distributed Learning
IRRODL

THE INTERNATIONAL REVIEW
OF RESEARCH IN OPEN AND
DISTRIBUTED LEARNING

Using Learning Analytics for Preserving Academic Integrity

Alexander Amigud, Joan Arnedo-Moreno, Thanasis Daradoumis and Ana-Elena Guerrero-Roldan

Cite this article

Article abstract

This paper presents the results of integrating learning analytics into the assessment process to enhance academic integrity in the e-learning environment. The goal of this research is to evaluate the computational-based approach to academic integrity. The machine-learning based framework learns students' patterns of language use from data, providing an accessible and non-invasive validation of student identities and student-produced content. To assess the performance of the proposed approach, we conducted a series of experiments using written assignments of graduate students. The proposed method yielded a mean accuracy of 93%, exceeding the baseline of human performance that yielded a mean accuracy rate of 12%. The results suggest a promising potential for developing automated tools that promote accountability and simplify the provision of academic integrity in the e-learning environment.

Using Learning Analytics for Preserving Academic Integrity

Alexander Amigud [1], **Joan Arnedo-Moreno**[1], **Thanasis Daradoumis**[1,2], and **Ana-Elena Guerrero-Roldan**[1]
[1] *Universitat Oberta de Catalunya (UOC), Barcelona, Spain,* [2]*University of Aegean, Mytilini, Greece*

Abstract

This paper presents the results of integrating learning analytics into the assessment process to enhance academic integrity in the e-learning environment. The goal of this research is to evaluate the computational-based approach to academic integrity. The machine-learning based framework learns students' patterns of language use from data, providing an accessible and non-invasive validation of student identities and student-produced content. To assess the performance of the proposed approach, we conducted a series of experiments using written assignments of graduate students. The proposed method yielded a mean accuracy of 93%, exceeding the baseline of human performance that yielded a mean accuracy rate of 12%. The results suggest a promising potential for developing automated tools that promote accountability and simplify the provision of academic integrity in the e-learning environment.

Keywords: electronic assessment, learning analytics, academic integrity

Introduction

The expansion of e-learning in higher education has been well noted in the literature (Buzdar, Ali, & Tariq, 2016). The growing variety of Massive Open Online Course (MOOC) offerings (Salmon, Gregory, Lokuge Dona, & Ross, 2015) and their ambition to obtain a credit-bearing status (Blackmon, 2016) denotes just that. So does the emergence of the "post-traditional learner," who craves control over how, where, and when to acquire the knowledge (Bichsel, 2013). These trends present new challenges, particularly with respect to academic credibility, because unlike instructional approaches, pedagogies, learning technologies and delivery methods that evolve overtime, the values of academic integrity remain

impervious to change. The credibility and integrity of learning entails an imperative need to establish a relationship of trust between learners and instructors. It will always be important to know who the students are and to be able to verify authorship of their work.

Maintaining academic integrity becomes an increasingly challenging exercise as physical entities become represented by virtual aliases, when the class size increases, when students are geographically dispersed, and when the teaching and assessment roles become disaggregated. The traditional methods for ensuring the trust relationship stays intact are difficult to translate to learning environments where students and instructors are separated by the time and space gap, and use technology to communicate (Amigud, 2013). These methods stipulate how, when, and where the assessment activities take place and are, at least partly, responsible for the disparity in expectations and experiences of post-traditional learners. When applied to the e-learning context, these approaches negate the very premise of openness and convenience, let alone administrative and economic efficiency. How many human invigilators are needed to ensure that all 10,000 learners enrolled in a course are not cheating? What does it cost? Who is paying the bill? What are the impacts on accessibility? These questions highlight a need for an effective, efficient, and robust academic integrity strategy. This approach must be capable of promoting accessibility, openness, and convenience, while allowing the natural evolution of e-learning towards a more accessible and open state.

This issue is timely and important because without an accessible, effective and efficient mechanism for mapping learner identities with the work they do across programs, courses, and individual assessments, institutions and e-learning providers run the risk of issuing course credits to anyone, simply by virtue of participation. Such a result inevitably affects institutional credibility.

In this article we present an analytics-based approach for aligning learner identities with the work they do in the academic environment. Unlike the traditional methods that rely on humans and/or technology to first, verify learner identities, and second, collect evidence to refute authorship claims, the proposed approach can concurrently verify learner identity and authorship claim. Therefore, by minimizing the number of verification steps, the proposed approach aims to provide greater efficiency, convenience, and accessibility. The contributions of this study are twofold: (a) An analytics-based academic integrity strategy that aligns learner identities with academic artifacts in a one-tiered approach; (b) A baseline of human performance for classifying student writings by author.

The research questions this article will answer are as follows:

- Is it possible to map student identities to their work through pattern analysis of the student-produced content?

- What is the performance level of the proposed approach?

- What is the performance baseline for instructor-based validation?

- What are the implications for practice and future research?

The article continues with an overview of related works, introduces the main theoretical concepts, discusses the study design and methods, and presents the results. The article concludes with a discussion of the implications for learning, teaching, administration, and directions for future research.

Background

Throughout the learning cycle, students produce academic content such as research reports, computer code, portfolios, and forum postings, which serve as the basis for performance evaluation and subsequent credit issuance. But how can instructors be sure that their students are not cheating? This question is not an easy one to answer and has been on academic administrators' radar for over two decades (Amigud, 2013; Crawford & Rudy, 2003; Grajek, 2016; Moore & Kearsley, 1996). The challenge stems from the two-tiered nature of academic integrity, comprising identity verification and validation of authorship processes. In other words, one needs confidence in knowing that the students are who they say they are, and that they did the work they claim to have completed. However, confidence comes with a cost and strategies that provide both the identity and authorship assurance are generally resource-intensive and invasive. As such, academic integrity is not delivered at a uniform level across all learning activities, but often applied selectively. This approach creates blind spots. For example, assignments submitted electronically may undergo plagiarism screening, but do not require identity verification. For example, online discussions are generally left unscrutinized, whereas the high-stakes final exams are often proctored and prior to entering the exam room students are required to present proof of identity. Academic integrity strategies can be classified into three types: (a) those that aim to verify student identities, (b) those that validate authorship claims, and (c) those that monitor and control the learning environment. They are summarized in Table 1.The effectiveness of academic integrity strategies is underexplored in the literature. There is also little discourse on the expected levels of performance that allows for comparisons to be drawn. Many of the studies focus on the perceptions of the students and faculty. This gap hinders the ability of instructors and academic administrators to make informed decisions for selecting strategies based on more than just the costs and the sentiment. Aside from efficiency issues, validating the authorship of every academic artifact is a resource-intensive task and the growing class sizes may only exacerbate the challenge of preserving academic integrity.

Table 1

Summary of Academic Integrity Strategies

Type	Method	Advantages	Disadvantages	Reference
Identity assurance	Biometrics	Provides high level of identity assurance and verification can be automated.	May require special hardware.	Apampa, Wills, and Argles (2010)
	Challenge questions	Accessible and convenient.	May be perceived as disruptive when	Bailie and Jortberg

		Verification can be automated.	employed continuously.	(2009)
Authorship assurance	Plagiarism detection	Automated reporting. Accessible and convenient.	The method is not designed to validate authorship but to dispute authorship claims.	Fiedler and Kaner (2010)
	Instructor validation	Supports continuous assessment. Reinforces the values of trust and integrity.	Resource intensive. Not readily scalable.	Barnes and Paris (2013)
Monitoring and control	Proctoring	Suitable for any assessment task. Eliminates travel requirements if conducted remotely. Can be automated.	Resource intensive. May affect accessibility due to scheduling and travel requirements. The method is not designed to validate authorship but to dispute authorship claims.	Li, Chang, Yuan, and Hauptmann, (2015).
	Activity monitoring	Accessible and convenient. Can be automated.	The method is not designed to validate authorship but to dispute authorship claims.	Gao (2012)

Many of the academic integrity approaches aim to collect evidence that may disqualify the assessment results. In the absence of such evidence, student-produced content is considered true and original. Evidence is collected through a variety of means ranging from direct observation to electronic communication.

Learning analytics and educational data mining have received a great deal of attention in recent years. The latter is concerned with methods for exploring educational data, while the former is concerned with measurement and reporting of events from data (Siemens & Baker, 2012). Some scholars introduced a notion of learner profiling, "a process of gathering and analyzing details of individual student interactions in online learning activities" (Johnson et al., 2016, p. 38). Students interact with learning technology, content, peers, and instructors (Moore & Kearsley, 1996). These interactions manifest in data that can be mined to provide the basis for decisions on enhancing learning experience, improving instructional design practices, and addressing security issues. Much of the analytical work is performed using machine-learning (ML) techniques (Domingos, 2012). ML is a growing field of computer science comprising computational theories and techniques for discovering patterns in data. A key distinction between ML and statistical techniques familiar to quantitative researchers is that ML algorithms learn from data without being programmed for each task.

Analytics has proven useful in the context of course security and integrity. The main benefit of using analytics for enhancing identity and authorship assurance is automation. Computational methods enable concurrent data analysis in the background, while the learner and instructor are actively engaged in learning and teaching. For example, keystroke analysis could be employed to validate learner identity (Barnes & Paris, 2013). A video stream of learners taking an exam can be analyzed to identify variations in environmental conditions such as the presence of other people in a room, or activities that may be restricted during a high-stakes assessment (O'Reilly & Creagh, 2015). Detection of plagiarism in written assignments employs algorithms that measure similarity in textual data (Kermek & Novak, 2016) and may automatically flag cases of plagiarism as papers are uploaded to the learning management system (LMS).

Content Analysis

The present study seeks to take advantage of the available learner-generated content and a habitual nature of language use (Brennan, Afroz, & Greenstadt, 2012; Brocardo, Traore, Saad, & Woungang, 2013). Writing style may be considered a form of behavioural biometrics (Brennan et al., 2012; Brocardo et al., 2013) and some have drawn an analogy to a fingerprint (Iqbal, Binsalleeh, Fung, & Debbabi, 2013). Writing patterns may be analyzed at the individual level, identifying features that are specific to a particular person. They can also be analyzed at the group level, classifying authors by age and personality type. Authorship analysis traditionally used for the resolution of literary disputes is now found to be useful for solving pragmatic issues such as forensic inquiries, plagiarism detection, and various forms of social misconduct (Stamatatos, Potthast, Rangel, Rosso, & Stein, 2015).

At least two studies examined performance of machine learning algorithms using a corpus of academic writings for the purpose of student authentication (Amigud, Arnedo-Moreno, Daradoumis, & Guerrero-Roldanm, 2016; Monaco, Stewart, Cha, & Tappert, 2013). A study by Amigud et al., 2016 analyzed a corpus of academic assignments and forum messages of 250 to 4,500 words by 11 students, using the Multinomial Naive Bayes algorithm and lexical features (word n-grams). The experiment yielded accuracy rates of 18%-100%. Inconsistency of topics, sample imbalance, and noise were some of the factors affecting the performance. In another study, Monaco et al., (2013) compared the performance of keystroke dynamics and stylometric authentication with a group of 30 university students. The results of the authorship analysis were inferior to that of keystroke dynamics-based analysis, although experiments yielded the performance rate of 74%-78% using a set of character, lexical, and syntactic features using the k-Nearest-Neighbor algorithm on texts ranging between 433 and 1,831 words.

In order to discriminate authorship styles, textual data need to be broken down into measurable units that represent writing behaviour, which are then extracted from the text and quantified. Style markers are often classified into five types (Stamatatos, 2009), although variation in nomenclature exists. These include: (a) character, (b) lexical, (c) syntactic, (d) semantic, and (e) application specific features. Authorship research has yielded approximately 1,000 stylistic features (Brocardo et al., 2013); however, the debate around the most effective set of features is still ongoing. For example, a study by Ali, Hindi, &Yampolskiy (2011) employed a set of 21 lexical, application specific features. Another study by Argamon, Koppel, Pennebaker, and Schler (2009) used a set of function words and the parts-of-speech as well as 1,000 content-based frequent words. N-gram-based features used in information retrieval tasks are also used in authorial discrimination tasks and vary in the type of units they represent. Some researchers have

successfully used character n-gram features (Brocardo et al., 2013) and others have proposed new variants using parts of speech tags, called syntactic n-grams or sn-grams (Sidorov, Velasquez, Stamatatos, Gelbukh, & Chanona-Hernández, 2014). For more information on trends in authorship analysis, please refer to the PAN/CLEF evaluation lab (Stamatatos et al., 2015).

Much of the research is aimed at discovering the right combination of style markers and computational techniques. Some of the common algorithms for text classification include: Support vector machine (SVM), Bayesian-based algorithms, Tree-based algorithms, and Neural Networks, among others (Aggarwal & Zhai, 2012). Multiple algorithms can also be used together to build ensembles. For example, predictions from multiple algorithms can be passed on to a simple voting function that selects the most frequent class. Ensemble methods often perform better than individual algorithms (Kim, Street, & Menczer, 2006; Raschka, 2015). However, classification performance depends on more than just an algorithm. There are a number of factors such as data set size, feature set size and type, proportion of training to testing documents, number of candidate authors, normalization technique used, and classifier parameters, among others that may influence the quality of predictions. The accuracy measure is often computed to report classification performance, which can be expressed as the number of correct predictions (true positives [TP] true +negatives [TN]) divided by the total number of predictions (true positives [TP], true negatives[TN], false positives [FP], and false negatives [FN]). More formally: ACC= (TP+TN)/(FP+FN+TP+TN).Results reported in the literature can vary widely, but most of the results fall within the 70%-100% range (Monaco et al., 2013).

Proposed Method

This research is part of our ongoing work to empower instructors with automated tools that promote accountability and academic integrity, while providing convenient, accessible, and non-invasive validation of student work. The rationale behind our method is that student-generated content is readily available and carries individual-specific patterns. Students employ perceptual filters and given the same input, different students respond differently. The comprehension of reality is conducted through the lens of existing beliefs and assumptions that impose limitations on how the perceived inputs are construed. This results in the production of academic artifacts containing a distinct signature, particular to each student. Therefore, artifacts produced by the same student are expected to be more similar to each other than to the work of other students. This allows a delineation of student-produced artifacts by analyzing stylistic choices exercised by students. These artifacts are then collected and stored for analysis, bearing a label of authorship that allows subsequent identification. The process of content creation is depicted in Figure 1.
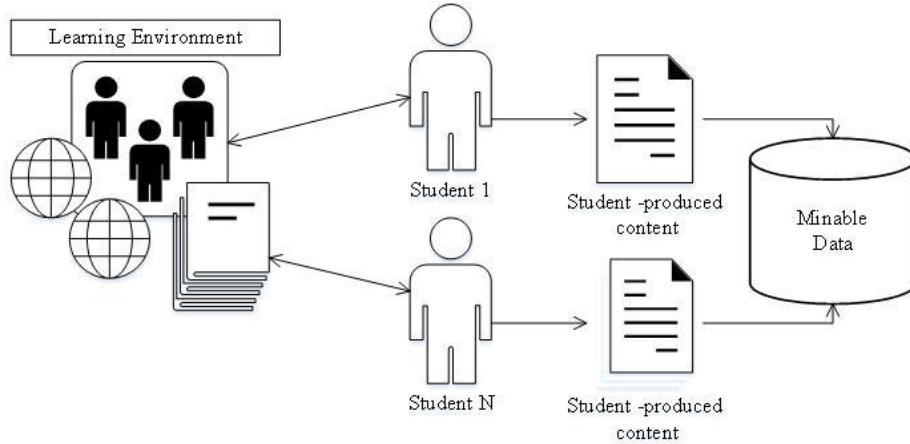
Figure 1. Content creation process.

Building upon previous research, this study aims to further enhance the automation of validation of student works by enabling machine-learning algorithms to select the most relative features of authorial style as well as using multiple algorithms in ensemble to attain better accuracy of predictions. The emphasis is on building an application that, provided with student-generated content, will produce a report highlighting areas that require the instructor's attention.

Because learning is a continuous process, content is readily available and can be mined throughout courses and programs. Machine learning provides the necessary means to learn from, and make guided decisions out of seemingly meaningless data. Prior academic work can be used to create a stylistic profile that will be tested against all subsequent student-produced content. Even before program enrollment, schools often require their prospective students to complete entrance exams, which may be used as inputs to the validation process of subsequent learning activities. The process of analyzing student assignments is depicted in Figure 2.
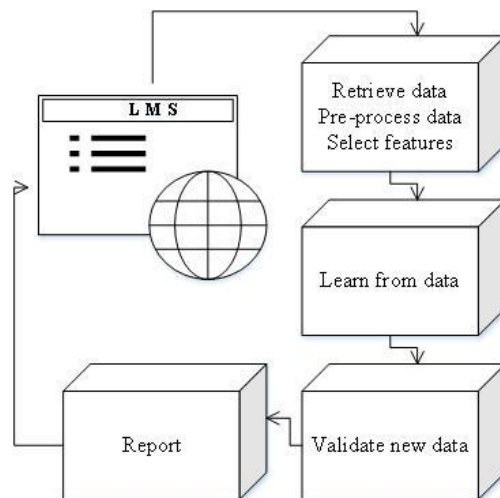
Figure 2. Data analysis of student-produced content.

The problem of aligning student identities with the work they do is posed as a classification task. Given a set of documents, an algorithm associates textual features with the identities of the students who produced them. When a new document is presented, the algorithm attempts to find a student whose use of textual features is more similar to the ones learned earlier. The student-generated content is passed on to classification algorithm(s) that learn to associate labels (student names) and patterns of language use from the examples in the training set and predict a class label which represents the students for each sample in the testing set. This prediction is then compared to the student names at the time of the assignment submission and any discrepancy in the predicted labels versus the student-supplied labels raises a red flag. To cover any blind spots in the academic environment, any cases of misclassification should be randomly examined by the instructor to ensure that the standards of academic integrity are maintained.

Figure 3 depicts two classification scenarios. Scenario A features an artifact claimed by Student X that was classified to be produced by Student X, and an artifact claimed by Student Y that was predicted to belong to Student Y. In contrast, Scenario B depicts a case of misclassification in which an artifact produced by Student Y bears similarity to the stylistic profile of Student X, in spite of being claimed by Y, which suggests a conflict and calls for the instructor's attention.
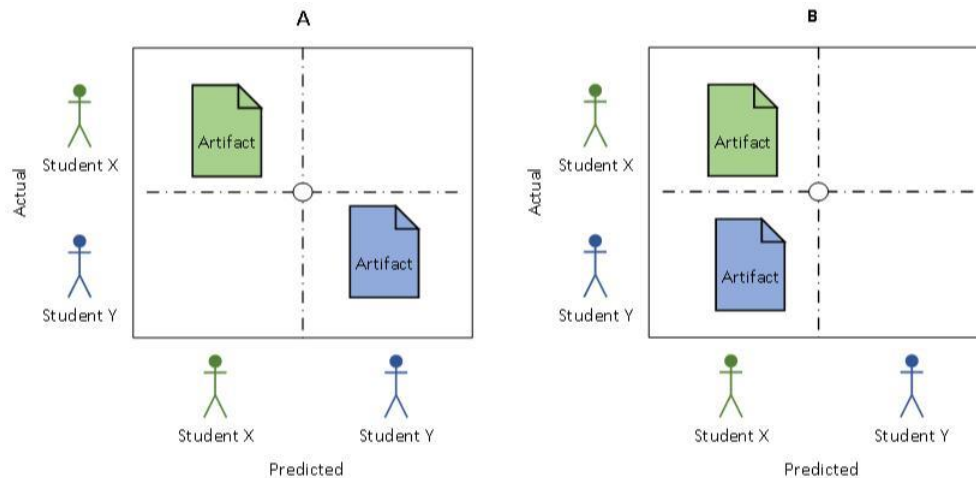

Figure 3. Classification scheme.

Study Design

To evaluate the proposed method, we have conducted two studies using a set of written assignments collected from two graduate-level research methods courses at an online European university. The first study examined the performance of computational techniques, and the second study examined the performance of the teaching staff. The second study helped to bridge the gap in the literature by providing a comparative performance baseline. In the next sections we describe data collection and analysis steps.

     Data set. In order to maximize the validity we employed real-world data. Our data set was composed of a subset of a corpus of student writings and included twenty written assignments by five students enrolled in two graduate-level research methods courses at an online European university. We employed a convenience sample with three inclusion criteria: (a) courses were delivered in English, (b) each course had at least one written assignment, and (c) participants were enrolled and successfully completed both courses. The courses employed an authentic assessment method. These assessments were low stakes and not proctored. The course integrity was maintained through instructor validation and an academic integrity policy. All of the students were non-native English speakers. After each learning activity assignments were uploaded by the student to the LMS for grading and then retrieved from the LMS document store for the analysis.

Our study was designed around the existing course formats, and all content was produced using tools and methods that students deemed fit in the circumstances. Students work with a variety of document formats including: Latex, portable document format (PDF), and Word documents. Students have used a variety of templates and formatting styles to present their work, hence there were differences in the amount of footnotes, headers, footers, in-text citations, bibliographies, and the amount of information on the front page.

Upon retrieval from the LMS datastore, the data were made anonymous and all identifiable information was replaced by a participant number. The corpus was composed of four assignments, two in each course, ranging between 1000 and 6000 words. The documents have undergone pre-processing steps, and noise - contributing items, (e.g., students' names, course numbers, and citations) were removed using a set of regular expression rules, fine-tuned with each iterative step to address specifics of the documents. Documents were split into chunks of 500 words. The collected data were assumed to be the ground truth; that is, the authorship claim for each document was considered to be true and that students were responsible for producing their own work while adhering to the academic standards.

In this study we used a set of lexical features comprising spanning intervening bigrams (non-contiguous word pairs) with the windows size =5, frequency of occurrence ≥ 3. Function words were preserved. In contrast, the traditional bigrams are contiguous. Each bigram is counted as a single token. Feature extraction was performed using the NLTK library (Bird, 2006). We also employed a syntactic set of 41 parts of speech (POS) tags extracted using the NLTK POS tagger; please see The Penn Treebank tag set for more information (Marcus, Marcinkiewicz, & Santorini, 1993).

Vectorization, or the process of creating feature vectors, transforms the human-readable textual elements into numerical values. For example, if co-occurring word pairs (bigrams) are defined as features, we count how many times each unique pair, counted as a single token, occurs in a given text. A weighting scheme is then applied to normalize the data. Both training and testing sets underwent the same pre-processing steps. For the lexical features, Term Frequency-Inverse Document Frequency (TF-IDF) weights were computed (Ramos, 2003). POS features were normalized by dividing by the number of tokens in the document. Some features are more important than others, and isolating them often helps to improve computational efficiency and performance. Considering the modest size of the training set, feature

selection using an Extra Trees Classifier was performed and the top 300 features were selected. Figure 4 depicts the processes of converting student assignments into machine readable format for classification.

Data mining techniques, although often used in the context of big data, may also prove useful in small data sets. Much of the authorship studies employ data sets with a small number of candidate authors or documents. For example, Sidorov et al. (2014) conducted an authorship attribution study using a set of three authors of 39 literary texts. In another study, Ali et al. (2011) examined methods for identification of chat bots using a set of 11 candidate authors. Even though the authorial pool or the corpus size is considered small, the feature space may be measured in thousands, because feature extraction techniques are only limited by computational capacity and the researcher's approach. The method of dimensionality reduction is often applied to eliminate noise and the possibility of over-fitting.
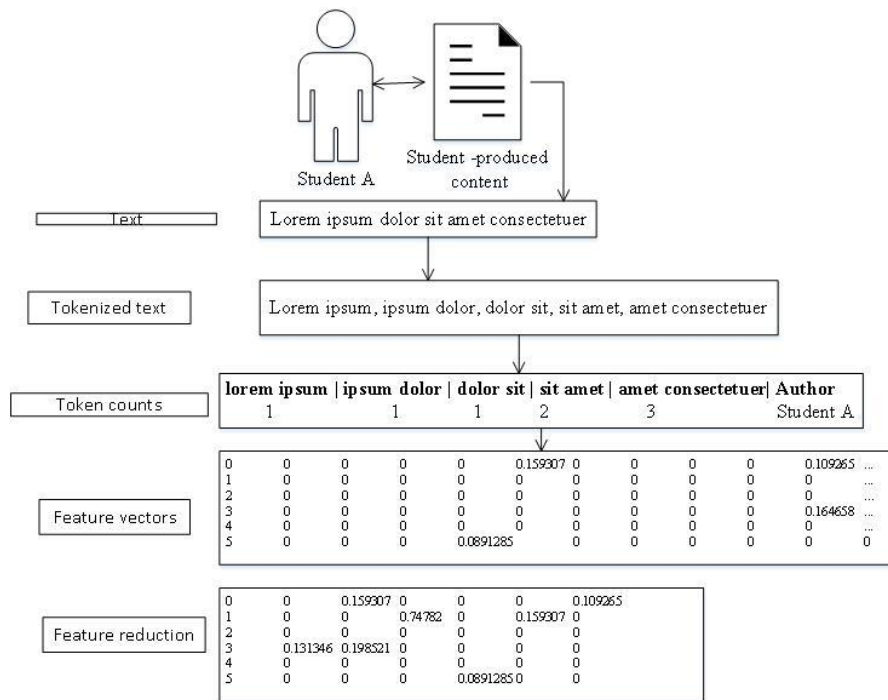

Figure 4. Style quantification process.

Data analysis. All analyses were performed using Python programming language and Scikit–Learn machine learning library, a set of tools for data mining and analysis (Pedregosa et al., 2011). Each student was defined as a unique class, where all writings of a single student shared a class label. From each document, features were extracted, normalized, weighted in order of importance and selected, then together with class labels passed on to an algorithm for training. A model was built from the patterns in data. The algorithm was provided with a new set of documents, this time without the labels to predict the student who authored each of the assignments in the test set, based on the patterns learned earlier. The experimental protocol was as follows:

- Retrieve student produced content (raw data)

- Process data (remove noise, split into chunks)

- Extract features

- Compute the feature importance

- Reduce feature space to the top 300 features

- Split data into training and testing sets

- Train classifier on training set

- Test classifier on testing set using each of the three algorithms

- Perform voting

- Report accuracy

This method differs from much of the literature in that not one, but three algorithms were employed using an ensemble method of majority rule voting (Raschka, 2015). Another distinction is that the labels of the test set are also known, because student- produced content is not anonymous. The analyses were conducted using: (a) c-Support Vector Classification (SVC) algorithm—Support Vector Machines (SVM) classifiers have been successfully used in a number of authorship studies (Sidorov et al., 2014); (b) Multinomial Naive Bayes another common algorithm for text classification tasks (Feng, Wang, Sun, & Zhang, 2016); and (c) Decision Tree classifier (Afroz, Brennan, & Greenstadt, 2012). Predictions by these classifiers are used as inputs to the majority voting algorithm, which measures central tendency (mode) of the predicted label for a sample by each of the classifiers. The most frequent class label wins. For example, if there are three classifiers in ensemble and two classes, Student A and Student B, and two of the classifiers predict that the assignment belongs to Student A, then the majority wins and the class is predicted as Student A. Figure5 demonstrates this example.
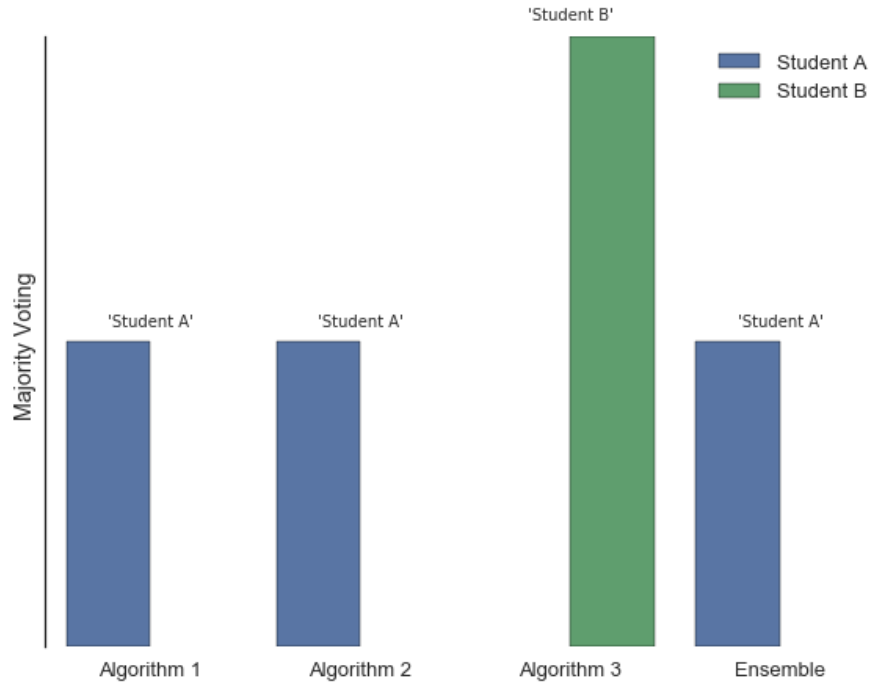
Figure 5. Majority voting example.

For more accurate examinations of the approach, experimental evaluation was conducted using the 10-fold cross-validation method, and used the same train/test split ratio as many of the other authorship studies (Brocardo, Traore, & Woungang, 2015; Schmid, Iqbal, & Fung, 2015). Data were randomly split into training and testing sets where 90% of the data were allocated for training and the remaining 10% allocated for testing. The accuracy measure was used to quantify the level of performance.

Human performance. The literature is sparse on the performance and effectiveness of academic integrity strategies. To bridge this information gap, we have conducted an additional study to establish a baseline of human performance and measure how well the practitioners directly responsible for grading assignments can identify patterns in students' writings. Barnes and Paris (2013) argued that the instructors should be able to identify instances of cheating or plagiarism once they become familiar with the style of a student's writing. We put this theory to the test. To our knowledge, this study is the first of its kind to compare the performance of academic practitioners to that of technology-based methods.

The experimental protocol was as follows:

1. Retrieve student assignments (raw data),

2. Select texts by five authors,

3. Process data (e.g., remove noise),

4. Create six question sets (5+ control),

5. Invite participants to perform classification, and

6. Compute accuracy of predictions.

The texts (500 word excerpts) were randomly distributed over five classification tasks. There were three texts per task. Two of the three texts were written by the same student. Five multiple choice questions accompanied each task, asking the participants to identify which texts were written by the same student. One task was used as a control, where all three texts were the same. Participants were included in the study if (a) they taught at an accredited university, (b) their professional responsibilities involved grading student assignments in the recent academic year, and (c) they correctly identified that the texts in the control task were by the same author. This study was conducted as part of a larger research project to examine instructor performance. We were using data from completed responses by 23 participants. The mean accuracy was calculated for the group.

Results and Discussion

Our proposed approach takes advantage of the available student-produced content and computational techniques that discover patterns in seemingly meaningless data in order to map learner identities with their academic artifacts. The experiments were conducted using real world data—a corpus of student assignments collected from two research methods courses. The results summary is depicted in Figure 6.
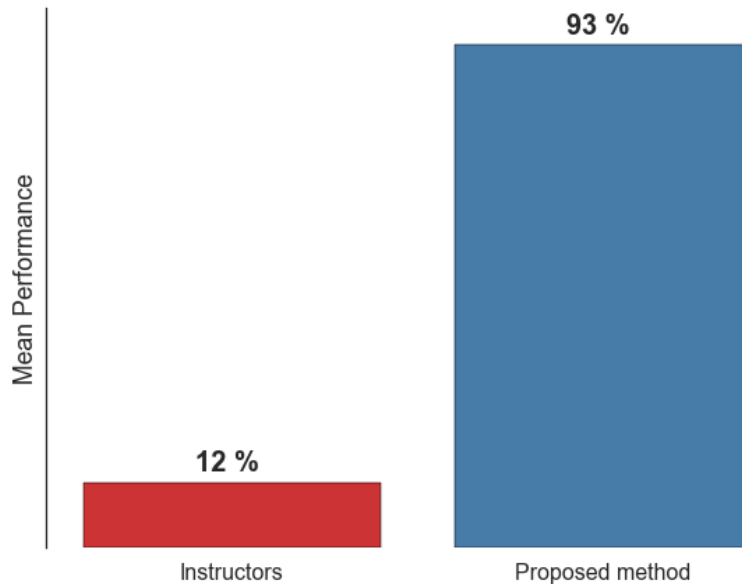

Figure 6. Prediction accuracy.

The ensemble method was able to map student-produced content with 93% accuracy. The method was tested over 10 independent train-test runs, where 10% of the data were randomly withheld, and the

classifier was trained on the remaining 90% of the data set. The results suggest an improvement over the earlier work on analytics-based academic integrity approaches.

Although higher accuracy is generally more desirable, the performance value in itself is not particularly meaningful without a comparative baseline. To this end we conducted a study to establish performance of the education practitioners classifying short academic texts. We examined Barnes and Paris's (2013) notion that instructors should be able to delineate students' writing styles. The results, however, suggest that the instructors performed at a significantly lower rate of 12%.

This study posed four research questions focusing on the performance of the proposed method to align learner identities with the academic work they do through pattern analysis. The first three questions were answered in the previous sections: in the following paragraphs we will address the last question about the implications for practice and future research.

The results suggest that the proposed approach to academic integrity can identify learners and map their artifacts with accuracy higher than that exhibited by teaching staff. The size and the type of data employed in these experiments is a limitation, and so are the feature sets and the algorithms used. Therefore, the results should be considered preliminary and further research is required to assess the performance of the proposed technique using a variety of data of different sizes and types. Although the human performance data provides a general idea of how well instructor-based validation approach detects inconsistencies or similarities in student writings, performance is expected to decrease in real-world settings with increases in the sample size, text size and type, fatigue, and other factors. This should be considered a limitation of the baseline estimate.

Nevertheless, the findings suggest a promising new avenue for addressing the issue of academic integrity through analytics, which has implications for future research, policy, and development of learning technologies. Data mining and analytics may constitute an alternative to the current observation-based strategies and provide a robust, non-intrusive and privacy-preserving method for obtaining information about who learners are and whether or not they did the work they claim they have done. The proposed approach holds the potential to eliminate the need to schedule invigilation sessions in advance, the need to employ human proctors, the need to share access to students' personal computers with third parties, or the need to monitor the student's environment via audio/video feed.

One of the key benefits of the computational-based approach is that validation of academic content can be performed in a continuous and automated fashion. This approach may be particularly valuable in courses that primarily employ authentic assessment methods (Boud & Falchikov, 2006; Mason, Pegler, & Weller, 2004; McLoughlin & Luca, 2002) that more often than not put low emphasis on identity and authorship assurance..The reporting function can also be automated and presented in a variety of formats such as e-mails, text messages and social media postings. Many institutions are starting to adopt some form of learning analytics, and adding another layer that targets student-content interaction will provide a single point of reporting, and expand the institutional capability to predict and mitigate risks to its credibility.

This approach can also be integrated with mobile learning services. It provides what other invigilation approaches cannot, which is validation of the authorship of student produced work as opposed to

gathering of evidence of academic misconduct. The proposed approach is not designed to completely relieve instructors from the burdens of maintaining course integrity, but rather to enhance their ability to detect incidents of cheating. The notion of performance has beenstressed throughout this paper and future research should examine the performance of other academic integrity strategies. The issue of academic integrity should be viewed holistically because the overall effectiveness of any academic integrity depends on more than just the technology, but requires sound policy, administrative, and pedagogical practices. Academic integrity and security are only as strong as the weakest link. The instructors will remain the first line of defense against cheating and it will be up to them to reinforce values, foster a culture of integrity and lead by example.

Conclusion

In this paper, we have proposed and examined experimentally a method of aligning student identities with the work they do by analyzing patterns in the student-generated content. To critically assess the relative performance of this approach, it was imperative to know the level at which human instructors are able to accurately classify student writings. To this end we conducted a study to measure the performance of teaching staff. The work described in this paper is part of a larger research program and in the future, the work will be expanded to obtain more precise measures of computational and human performance.

Contingent upon further research using larger and more diverse data sets, the proposed technology could find its way into the classroom. Analytics enables the automation of identity and authorship assurance, calling for an instructor only in cases where manual intervention is required. Furthermore, unlike the traditional academic integrity measures, the proposed method can continuously and concurrently validate the submitted academic work and provide both identity and authorship assurance. This may yield greater convenience, efficiency, openness, accessibility, and integrity in the assessment process. This also enables a greater level of user privacy as it eliminates the need to share access to students' personal computers or the need to share physical biometrics with a third party in order to complete identity verification (Levy, Ramim, Furnell, & Clarke, 2011). When computational techniques are used to align learner identities with their work, assessment activities become less intrusive to the learner and less logistically burdensome for the academic staff. Learning and institutional analytics is often used to track student learning and streamline organizational resources. Its application should be expanded to further promote the values of integrity and trust.

References

Afroz, S., Brennan, M., & Greenstadt, R. (2012). Detecting hoaxes, frauds, and deception in writing style online. *2012 IEEE Symposium on Security and Privacy.* doi:10.1109/sp.2012.34

Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. In C. C. Aggarwal, & C. Zhai (Eds.). *Mining text data* (pp. 163–222). New York: Springer. doi:10.1007/978-1-4614-3223-4_6

Ali, N., Hindi, M., & Yampolskiy, R. V. (2011). Evaluation of authorship attribution software on a Chat bot corpus. *2011 XXIII International Symposium on Information, Communication and Automation Technologies*, Sarajevo, 1–6.doi:10.1109/icat.2011.6102123

Amigud, A. (2013). Institutional level identity control strategies in the distance education environment: A survey of administrative staff. *The International Review of Research in Open and Distributed Learning, 14*(5).doi:10.19173/irrodl.v14i5.1541

Amigud, A., Arnedo-Moreno, J., Daradoumis, T., & Guerrero-Roldan, A. E. (2016). A behavioral biometrics based and machine learning aided framework for academic integrity in e-assessment. *2016 International Conference onIntelligent Networking and Collaborative Systems (INCoS)*, Ostrawva, Czech Republic, 255–262. IEEE.doi:10.1109/incos.2016.16

Apampa, K. M., Wills, G., & Argles, D. (2010). User security issues in summative e-assessment security. *International Journal of Digital Society (IJDS), 1*, 135–147. Retrieved from http://www.infonomics-society.org/IJDS/User%20Security%20Issues%20in%20Summative%20E-Assessment%20Security.pdf

Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2009). Automatically profiling the author of an anonymous text. *Communications of the ACM*, *52*(2), 119. doi:10.1145/1461928.1461959

Bailie, J. L., & Jortberg, M. A. (2009). Online learner authentication verifying the identity of online users. *MERLOT Journal of Online Learning and Teaching*, *5,* 197– 207. Retrieved from http://jolt.merlot.org/vol5no2/bailie_0609.pdf

Barnes, C., & Paris, L. (2013). *An analysis of academic integrity techniques used in online courses at a southern university.* Paper presented at the Northeast Region Decision Sciences Institute Annual Meeting,New York, NY. Retrieved from http://www.nedsi.org/proc/2013/proc/p121024007.pdf

Bichsel, J. (2013). The state of e-learning in higher education: An eye toward growth and increased access. *EDUCAUSE Center for Analysis and Research.* Retrieved from https://net.educause.edu/ir/library/pdf/ers1304/ers1304.pdf

Bird, S. (2006). NLTK. *Proceedings of the COLING/ACL on Interactive Presentation Sessions*, Sydney, Australia, 69–72. doi:10.3115/1225403.1225421

Blackmon, S. J. (2016). Through the MOOCing glass: Professors' perspectives on the future of MOOCs in Higher Education. *New Directions for Institutional Research*, *2015*(167), 87–101. doi:10.1002/ir.20156

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399–413. doi:10.1080/02602930600679050

Brennan, M., Afroz, S., & Greenstadt, R. (2012). Adversarial stylometry. *ACM Transactions on Information and System Security, 15*(3), 1–22. doi:10.1145/2382448.2382450

Brocardo, M. L., Traore, I., Saad, S., & Woungang, I. (2013). Authorship verification for short messages using stylometry. *2013 International Conference on Computer, Information and Telecommunication Systems (CITS).* doi:10.1109/cits.2013.6705711

Brocardo, M. L., Traore, I., & Woungang, I. (2015). Authorship verification of e-mail and tweet messages applied for continuous authentication. *Journal of Computer and System Sciences, 81*(8), 1429–1440. doi:10.1016/j.jcss.2014.12.019

Buzdar, M. A., Ali, A., & Tariq, R. U. H. (2016). Emotional intelligence as a determinant of readiness for online learning. *International Review of Research in Open and Distributed Learning, 17*(1). doi:10.19173/irrodl.v17i1.2149

Crawford, G., & Rudy, J. A. (2003). Fourth annual EDUCAUSE survey identifies current IT issues. *Educause Quarterly*, *26*(2), 12–27.Retrieved from http://net.educause.edu/ir/library/pdf/eqm0322.pdf

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78. doi:10.1145/2347736.2347755

Feng, G., Wang, H., Sun, T., & Zhang, L. (2016). A term frequency based weighting scheme using naïve Bayes for text classification. *Journal of Computation and Theoretical Nanoscience, 13*(1), 319–326. doi:10.1166/jctn.2016.4807

Fiedler, R., & Kaner, C. (2010). Plagiarism-detection services: How well do they actually perform? *IEEE Technology and Society Magazine, 29*(4), 37– 43. doi:10.1109/MTS.2010.939225

Gao, Q. (2012). Using IP addresses as assisting tools to identify collusions. *International Journal of Business, Humanities and Technology, 2*(1), 70– 75. Retrieved from http://www.ijbhtnet.com/journals/Vol_2_No_1_January_2012/8.pdf

Grajek, S. (2016). Top 10 IT issues, 2016: Divest, reinvest, and differentiate. *EDUCAUSE Review, 51*(1). Retrieved from http://er.educause.edu/articles/2016/1/top-10-it-issues-2016

Iqbal, F., Binsalleeh, H., Fung, B. C. M., & Debbabi, M. (2013). A unified data mining solution for authorship analysis in anonymous textual communications. *Information Sciences, 231*, 98–112. doi:10.1016/j.ins.2011.03.006

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC horizon report: 2016 Higher education edition*. Austin, Texas: The New Media Consortium. Retrieved from https://library.educause.edu/resources/2016/2/2016-horizon-report

Kermek, D., & Novak, M. (2016). Process model improvement for source code plagiarism detection in student programming assignments. *Informatics in Education-An International Journal, 15*(1), 103–126. doi:10.15388/infedu.2016.06

Kim, Y., Street, W. N., & Menczer, F. (2006). Optimal ensemble construction via meta-evolutionary ensembles. *Expert Systems with Applications*, *30*(4), 705–714. doi:10.1016/j.eswa.2005.07.030

Levy, Y., Ramim, M., Furnell, S., & Clarke, N. (2011). Comparing intentions to use university-provided vs vendor-provided multibiometric authentication in online exams. *Campus-Wide Information Systems, 28*, 102– 113. doi:10.1108/10650741111117806

Li, X., Chang, K., Yuan, Y., & Hauptmann, A. (2015). Massive open online proctor: protecting the credibility or MOOCs certificates. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW'15)*, Vancouver, 1129–1137. doi:10.1145/2675133.2675245

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The penn treebank. *Computational Linguistics*, *19*(2), 313–330. Retrieved from http://repository.upenn.edu/cgi/viewcontent.cgi?article=1246&context=cis_reports

Mason, R., Pegler, C., & Weller, M. (2004). E-portfolios: An assessment tool for online courses. *British Journal of Educational Technology, 35*(6), 717–727. doi:10.1111/j.1467-8535.2004.00429.x

McLoughlin, C., & Luca, J. (2002). A learner-centred approach to developing team skills through web-based learning and assessment. *British Journal of Educational Technology, 33*(5), 571–582. doi:10.1111/1467-8535.00292

Monaco, J. V., Stewart, J. C., Cha, S.-H., & Tappert, C. C. (2013). Behavioral biometric verification of student identity in online course assessment and authentication of authors in literary works. *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Arlington, VA, 1–8. doi:10.1109/btas.2013.6712743

Moore, M. G., & Kearsley, G. (1996). *Distance education: a systems view*. Boston, MA: Wadsworth.

O'Reilly, G., & Creagh, J. (2015). Does the shift to cloud delivery of courses compromise quality control. *Higher Education in Transformation Conference (HEIT)*, Dublin, Ireland, 417-424. Retrived from http://arrow.dit.ie/cgi/viewcontent.cgi?article=1001&context=st5

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830. Retrieved from http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.1424&rep=rep1&type=pdf

Raschka, S. (2015). *Python machine learning*. Birmingham, UK: Packt Publishing.

Salmon, G., Gregory, Lokuge, J., Dona, K., & Ross, B. (2015). Experiential online development for educators: The example of the Carpe Diem MOOC. *British Journal of Educational Technology, 46*(3), 542–556. doi:10.1111/bjet.12256

Schmid, M. R., Iqbal, F., & Fung, B. C. M. (2015). E-mail authorship attribution using customized associative classification. *Digital Investigation, 14*, S116–S126. doi:10.1016/j.diin.2015.05.012

Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., & Chanona-Hernández, L. (2014). Syntactic N-grams as machine learning features for natural language processing. *Expert Systems with Applications, 41*(3), 853–860. doi:10.1016/j.eswa.2013.08.015

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254). New York: ACM.doi:10.1145/2330601.2330661

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology, 60*(3), 538–556. doi:10.1002/asi.21001

Stamatatos, E., Potthast, M., Rangel, F., Rosso, P., & Stein, B. (2015). Overview of the PAN/CLEF 2015 evaluation lab. In J. Mothe, J. Savoy, J. Kamps, K. Pinel-Sauvagnat, G. Jones, E. San Juan, … & N. Ferro (Eds.). *Experimental IR meets multilinguality, multimodality, and interaction.* Springer. doi:10.1007/978-3-319-24027-5_49