An Initial Analysis Of The Contextual Information Available Within Auction Posts On Contract Cheating Agency Websites

Thomas Lancaster  
School of Computing, Telecommunications and Networks  
Birmingham City University  
Birmingham, United Kingdom  
thomas.lancaster@bcu.ac.uk

Robert Clarke  
School of Computing, Telecommunications and Networks  
Birmingham City University  
Birmingham, United Kingdom  
robert.clarke@bcu.ac.uk

Abstract — The advantages of using contextual information in order to detect contract cheating attempts by students have not yet been fully explored in the academic literature. Contract cheating occurs when a student uses a third party to produce assessed work for them. This paper focuses on contract cheating using agency websites, where an auction type process is used by students to select a contractor to have the assessed work produced for them, often at a financially advantageous price. Currently, the process of finding contract cheating on agency sites is manually intensive, with a detective required to investigate and attribute each cheating attempt. This paper aims to formally identify the context internally and externally available for contract cheating posts on an agency website. The paper is offered as a starting point for academics interested in producing an automated intelligent contextually-aware tool to detect contract cheating.

Keywords—contract cheating, plagiarism, student cheating, auction posts, agency web sites, academic integrity

I. INTRODUCTION

When academics institutions are making awards of qualifications to students, they need to be alert that the work being submitted may not be the student’s own.

The issue of student plagiarism has been comprehensively explored within the academic literature [1,2]. Plagiarism occurs when a student takes the words or ideas of another person and then hands this in as if it were their own work for academic credit. Traditionally, plagiarism sources might include textbooks or lecture notes, or the student may be colluding with one or more of their peers to submit closely related work.

The more modern literature on student plagiarism focuses primarily on the World Wide Web as being a typical source for unoriginal content [3,4,5]. This can make it as easy for a student to cheat using a simple “copy and paste” process.

Software solutions exist for many types of student plagiarism which seem to work successfully. Where students are working on developing source code programs, an area common within the Computing discipline, both of the freely available tools Measure of Software Similarity (MOSS) [6] and JPlag [7] have been shown to be effective at detecting similar work [8]. Simple metrics, such as looking for common pairs of consecutive source code tokens, have been shown to be sufficient to solve what is now considered a computationally simple task [9,10,11]. Solutions using visual approaches, designed to assess tutors in decided whether similarity represents plagiarism, are also available [12].

For purely textual submissions, such as student reports and essays, the commercial service Turnitin [13] is commonly used in higher education. Using this service has the advantage of comparing the student work to a vast international database of previous submissions, as well as its own archive of Web content. Other techniques using text matching tools have also been shown work well [14,15]. The simple technique of evaluating the proportion of consecutive word pairs, or bigrams, student submissions have in common has been shown to be largely suitable for free text [16,17], although limited gaps in the Turnitin database have been identified [18,19,20]. Alternative approaches to Turnitin using free software tools have also been identified [21].

There are still outstanding problems, where ways in which students can cheat on assessed work are known. These types of cheating are not easily detected using existing computing solutions. This is where it has been speculated that intelligent context-aware systems may prove to be invaluable [22].

At the top of the anti-cheating agenda is the issue of contract cheating [23,24]. This type of cheating occurs where a student has an original solution produced for them, which they then hand in as if they had completed this academic work for themselves.

Most often, contract cheating will take place when a student pays a third party or uses an online service to create the unoriginal work. This cheating may take a written essay form or may represent software source code. Within a wide-range of disciplines, this could also represent a creative component. For instance, the student could outsource the production of an original musical composition or a piece of artwork.

Contract cheating has also been observed to be produced without money changing hands. As an example, this could be created by a friend or family member wishing to share their skills to help a student to succeed. There could also be a barter system in place to let people take advantage of the different skill sets that exist inside a community. Trading an academic writing service, in return for gardening, for example, may be possible.
Whilst all of these methods of contract cheating show enterprise, a skill increasingly asked for by employers looking to recruit students ready for industry, they continue to represent a threat to academic integrity. Current software for detecting plagiarism will not find attempts to contract cheating. This software relies on being able to match work submitted by a student to a known source document. Since the work submitted as a result of contract cheating is original, these source documents will not be available. Therefore, contract cheating is unlikely to be detected using existing anti-plagiarism software.

Much of the current literature on contract cheating is focused on raising awareness and suggesting methods that academics can use to redesign assessment to make contract cheating untenable for students [24,25]. Whilst such publicity holds clear value, there have been few anti-contract cheating software solution proposed, and implementation attempts have not yet been successful.

Current methods of stopping contract cheating instead rely on a physical process that is labour-intensive for humans [26]. The role of the contract cheating detective has been identified [27]. This is a person who manually scans the content on websites where contract cheating attempts have been known to be posted. The detective then attempts to attribute the contract cheating by identifying the academic institution and course concerned, and, ideally, the particular student who is looking to outsource the production of a solution. The process is largely comparable with a forensic investigation, with the detective attempting to collect whatever digital evidence is made available. Due to the ad-hoc nature of monitoring, and the incompleteness of identifying information available to attribute contract cheating attempts, this process has been noted as being only partially successful [27].

Where an assignment specification is posted online, contextual information related to that assignment is available. An initial proposal related to using that context suggested that this could be linked to an international database of assignment specifications, which all staff contribute their assignments to [22]. That potential solution may be unattainable. Not all academic institutions appear to be willing to attest that contract cheating is a sector wide problem and so would not encourage their staff to contribute.

This paper instead focuses on the direct contextual evidence that can be collected from online contract cheating attempts, to provide the detective with a base point from which to attribute attempts at contract cheating. Such an evidence base can be considered the starting point for a fully intelligent context-aware software solution for contract cheating. Even a simple tool to automatically collect evidence should save the detective time, increase the comprehensiveness of the contextual data collected, and ensure that attempts at contract cheat are not missed.

The remainder of the paper is structured as follows:

- The main academic research relating to contract cheating is briefly reviewed, so as to provide a clear starting point for the contextual discussion.
- One particular type of online contract cheating, that is cheating using agency sites and auction sites, is investigated in detail for contextual information. An initial classification of the types of information available, both internal to the cheating attempt, and relying on intelligent gathering from external sources, is given.
- The longer-term applications of using this context are then considered. This is intended as the starting point for the provision of full tool support for this detecting contract cheating.

Whilst a complete solution to contract cheating does not yet appear possible, it is noted that advances in computing power and large-scale intelligent context-aware systems are opening up new opportunities for detection. Some initial proposals to a variety of potential contextual solutions for student cheating have been previously made [22]. It is intended that this paper can be used to form an early contribution to that continued wider discussion.

II. A BRIEF OVERVIEW OF CURRENT CONTRACT CHEATING RESEARCH

A. The Agency Site Process

Despite the apparent interest in contract cheating from articles in the media, this has not been regularly reflected by a large volume of academically styled publications.

The first use of the term contract cheating can be traced back to Clarke and Lancaster’s initial analysis of how students used the website RentACoder to cheat [23]. RentACoder has since been renamed as vWorker and merged into the larger outsourcing site Freelancer [28]. Sites of the type being discussed in this paper, where students attempt to outsource their work based on a process of bidding, are referred to as agency sites [29,30].

The focus of the initial study of RentACoder was on an auction like-process used by students to solicit the production of original work [23]. A similar process is still very much evident. The market leading agency site for contract cheating currently seems to be Freelancer [28], which has taken over several smaller sites, of which RentACoder is one example.

Freelancer is a site set up primarily to provide legitimate outsourcing services for small businesses. For instance, a company may use Freelancer to outsource copy writing for their web site. The concern comes when students use Freelancer for contract cheating. Since Freelancer does not monitor the auction requests that are made, student work is often posted. Freelancer does have processes that academics can use to get attempts to outsource academic work removed upon request.
The way the contract cheating auction process on agency sites works is as follows:

First, the student posts a request for an assignment to be completed. The complete details of the assignment specification may be available to the public. Other methods through which the assignment details can be publicised have been observed; these may include restricting the assignment specification to be sent through private messages when requested, or only making the assignment details available to specific geographical areas.

At that point, the bid request is available for people to make a financial offer to complete it. This is a competitive process, where workers quote the price that they’re willing to pay to complete the work for the student.

The student can then make their selection of workers from the offers presented. Some students may elect to take the lowest priced offer. Other students may review the feedback provided for the worker from their previous jobs, or discuss the assignment requirements in private in order to choose the worker who they believe will best produce the final assignment to the quality that is required.

Freelancer then takes an advance payment from the student, which is held in escrow. The worker is then instructed to begin the production of the assignment solution. Once work is produced that the student accepts, the payment is released to the worker, minus the commission kept by Freelancer for facilitating the process. Where work is not produced to the quality required, the student and worker can enter into a financial arbitration process.

At the end of the contract cheating process, the student has been provided with an original piece of work. Provided that they selected a genuine worker, who would not themselves take shortcuts, the finished work would appear as original as any other submitted for marking.

A study of the standard of programming assignments produced through contract cheating, by Jenkins and Helmore, found that these were economically priced and of sufficient quality to pass [3]. They also found that even blatant attempts of contract cheating positioned in front of marking staff were not detected. This suggests that some students taking computing courses are likely to be met with a large amount of results.

Attempts to outsource work at all levels, from introductory programming through to final year projects and dissertations, were identified.

Other research has focused on agency sites set up to provide assignment solutions for students, but without containing any commercial work [27]. One former example of this type of agency site was EssayBay [32]. The study of EssayBay provided evidence of 627 attempts by students to cheat and found subject areas outside of the usual Computing domain found on RentACoder. The subject areas Business and Administration Studies, Social Studies and History and Philosophical studies were seen to dominate the use of this site. The combined category, Mathematics and Computer Science, ranked only eighth in the frequency of use.

An investigation looking at the value of contract cheating and the extent of commercialisation of this problem found that over $1 million dollars of business per year runs through agency websites [30]. These sites provide enough information and context to be monitored by contract cheating detectives and so are the main sources used for current quantitative research into contract cheating. The $1 million total is not a complete estimate of the full commercial value of contract cheating, since there are many other essay writing services and personal tutoring services operating independently of the agency site model. There are also contract cheating agency sites which are private and not easily accessed by detectives, so they cannot be included in this estimate. An investigation for the trade publication, Times Higher Education, cited a claim that a single bespoke essay writing company “was worth £200 million a year” [33]. It seems likely that the real figure lies somewhere in between the two.

C. Possible Automated Solutions To Contract Cheating

Whilst possible solutions to contract cheating have been proposed in the academic literature, there do not yet seem to be any successful implementations. Instead, contract cheating detection relies on a physical human role, named as the detective, who manually scans auction sites, attempts to identify the academic staff responsible for assignment specifications and then to notify the academics that one of their students appears to be cheating [26].

Previous studies have found that attributing assignments found online to staff is problematic, since most assignment details are kept inside private databases, and so not available to search [27]. A lack of visibility of staff contact details, such as email addresses, has also been noted. There can also be problems when a standard assignment specification is used, such as that taken from a textbook. Any attempts to match such an assignment specification to a host institution are likely to be met with a large amount of results.

Clarke and Lancaster proposed a Six-Stage Contract Cheating Process, intended to mirror the way this problem is approached by detectives in real life [26]. This process is shown in Fig. 1. The process would seem to be suitable to build upon with an intelligent context-aware system.
The process starts with a publication stage, where tutors publish copies of their assignment specification to a searchable database. More recent studies have suggested that Turnitin may be an appropriate database for this [22,30]. Any request submitted to an agency site is automatically stored for evidence and analysed in the Collection stage. The Identification stage sees a detective provide human confirmation whether the request is legitimate commercial work, or if they are a student wanting to outsource an assignment. The Attribution stage sees the correct academic institution for the work identified using the original database as an aid. The Notification stage sees an appropriate academic being told that their assignment has been found online. Finally, the Investigation stage, which is outside of the detective’s immediate role, sees the miscreant student being identified and an appropriate academic misconduct process being followed.

An appropriate intelligent system would reduce the load on the detective, particularly during Identification and Attribution and would ensure that a fair and consistent process for finding student posts on auction sites is always followed. The need to automatically collect content and to build up databases of assignment specifications and agency site content, was previously identified by Lancaster and it would seem appropriate for those databases to be contextual in nature[22].

An alternative automated detection approach that has been identified would use stylometrics to identify if a solution handed in by a student appeared to have been written by another person [34]. Initial work by Culwin pondered whether it would be possible to track the writing style of a student across a number of assignments. This would allow tutors to identify when the writing style of one assignment differed substantially from those submitted by the same student previously. A simple measure with which this could be done would be the document reading age, for example.

The current tests results of using stylometrics to detect contract cheating are inconclusive, but this is only an emerging field of research. It is hoped that an intelligent system, incorporating machine learning capability to more comprehensively understand how a student writes assessed work, may be able to overcome the current challenges with using stylometrics.

An alternative automated detection approach that has been identified would use stylometrics to identify if a solution handed in by a student appeared to have been written by another person [34]. Initial work by Culwin pondered whether it would be possible to track the writing style of a student across a number of assignments. This would allow tutors to identify when the writing style of one assignment differed substantially from those submitted by the same student previously. A simple measure with which this could be done would be the document reading age, for example.

The current tests results of using stylometrics to detect contract cheating are inconclusive, but this is only an emerging field of research. It is hoped that an intelligent system, incorporating machine learning capability to more comprehensively understand how a student writes assessed work, may be able to overcome the current challenges with using stylometrics.

A. An Illustrative Example Of An Attempt At Contract Cheating Found On An Agency Site

Fig. 2 is indicative of the style of Computing assignment observed by detectives on Freelancer.com. It is a request to have a final dissertation written for an Information Systems course, including a tool for search engine optimization analysis and the associated report [35]. In this case, the attempt to outsource the dissertation does not appear to have been successful, although the student may have made alternative arrangements.

The closed post on the agency site shows that 14 bids were made, with an average of $575. Since the student has specified a 60 page written report, this equates to a low $9.58 per page. On top of this, for no additional income, the successful worker would need to produce the necessary software. Limited details of the poster, the requested assignment and the workers who have bid to produce the dissertation are available.

This section identifies the contextual information available as part of a typical bid like this, as would be used by a human detective. This context could then be used by an intelligent system attempting to solve contract cheating, in conjunction with a standard process like the Six-Stage Contract Cheating Detection Process [26]

B. Context Available Directly On The Agency Site

The following context relating to the student who has posted the request for work may be available.

- The username provided by the student, which may match to usernames on other sites
- The stated location provided by the student. This location may not be accurate
- The photo provided by the student, which may match to photos on other sites
- The text provided with the bid, which may indicate the source of the assignment
• Attachments relating to the assignment specification. These may directly contain details of the source of the assignment, or they may contain trace evidence, for instance, when the student has attempted to remove the name of the academic institution.

• Identifiers left with the assignment. These might include module codes, locations, staff names or learning outcomes.

• The choice of language for the assignment. For instance, UK English may indicate that the worker is studying at a UK university.

• Previous jobs placed by the student, which may also contain valuable contextual information.

• Comments provided by workers involved with the project, for instance in the form of feedback, which may give away identifying information.

   It is important that all this information is collected as soon as a request for work goes live, since requests may be removed or hidden at any time.

C. Context Available Externally To The Agency Site

For successful attribution and notification, an intelligent context-aware system would need access to as much of this contextual information as possible:

• A database of all assignment specifications issued worldwide. This may be stored on Turnitin [13].

• A database of all student projects, dissertations and individualised assessments.

• An accompanying database of contact information for all of these, to enable notification.

• Access to social media profiles for students, or an approximation, as provided by searches of Google and Google Images.

Where databases are not available, it may be possible to approximate these, so long as assessment details are available through Google.

The challenge to the detective is that successful attribution often requires the use of a number of indicators and is a problem of missing information. For instance, a stated location may be used to identify the vicinity in which a student is based, with the assumption made that they attend a local university. For an intelligent system looking to detect contract cheating to be successful, it is likely that a large training data set would be necessary. Having access to such a data set may also allow for advances in stylometrics techniques applied to contract cheating [34].

IV. CONCLUSIONS AND RECOMMENDATIONS

Contextual information is available to support developers wishing to investigate automatic methods for detecting contract cheating using intelligent systems. However, these are dependent on the external improvements to setting assignments which also represent good practice, such as ensuring that all assignment specifications are original and traceable.

A further complication awaits any academic receiving details that their assignment specification has been found online. In some cases, it may be possible to identify the student from the information provided, however, observations from detectives suggest that this is only possible with about one in three attributed assignments [30]. With a large class, with several hundred students, it may be impossible for an academic to pinpoint the at-fault student.

The method of individualising assessments has been proposed [36]. Here, each student has a unique assignment specification assigned to them. For instance, they may receive a unique subset of five software requirements out of eight. If this assignment specification is found online, the unique combination can be used to identify the cheating student, even if embedded names have been removed.

Observation suggests that there is room for a toolkit and educational advice to aid tutors with creating sensible individualised assignments. This would be particularly useful outside the Computing discipline, where staff may lack the technical development skills needed to produce individualised assignment specifications. A toolkit could also embed invisible watermarks, such as a unique combination of spaces, to further help identification. All individualised assignments would need to be submitted to the database being established to support the intelligent context-aware contract cheating detection system.

The field of stylometrics also seems to show promise as the basis for such a system [34]. One of the possible issues would seem to be ensuring that all the work used to train the system was created by the student listed, and not acquired elsewhere. Other studies have observed some success with tracking student typing patterns for online tests and looking for frequent misspelled words [37]. This may suggest a sensible approach for a new system to take.

An intelligent detection system for contract cheating cannot be the only answer. Continued training of tutors and vigilance by them is needed to ensure that cheats do not prosper. One development observed by their authors in their detective work has been the appearance of trace elements in similarity reports obtained from Turnitin. Here a small match of 1% can be substantial if it contains the name of a recognisable essay writing or agency site. A system relying on Turnitin to speed up the development of a contract cheating detection system would need to be aware of the value of those small matches. Even when tutors are doing routine checks of student work for originality, this suggests that they need to check all student submissions, not just those obtaining high similarity scores.
Contract cheating looks likely to be a growing problem when students take the assessments associated with Massive Open Online Courses (MOOCs). Considerable issues have already been observed with attribution of posted assignments for courses offered at universities that solely operate online [27]. With the sophistication of the commercial world of contract cheating, it is likely that workers will spring up specialising in the assignments set for specific MOOCs. That growing sector will need to ensure that they have anti contract cheating processes in place for their qualifications to hold any value.

For face-to-face courses, traditional methods of assessment should not be neglected. For instance, a practical programming exam may solve the problem of an outsourced assignment [24]. A spoken test on a submitted essay may allow a tutor to spot whether or not this is the student’s own work. There may be the possibility of an intelligent system to simulate what would otherwise be a labour intensive testing process for a human. A combination of these additional techniques may finally allow the contract cheating challenge to be solved.

REFERENCES