Forensic Science International 297 (2019) e8–e10

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Short Communication

Chat Analysis Triage Tool: Differentiating contact-driven vs. fantasy-driven child sex offenders

Kathryn C. Seigfried-Spellar*, Marcus K. Rogers, Julia T. Rayz, Shih-Feng Yang, Kanishka Misra, Tatiana Ringenberg

Purdue University, West Lafayette, IN, United States

ARTICLE INFO

Article history:
Received 27 July 2018
Received in revised form 6 February 2019
Accepted 7 February 2019
Available online 23 February 2019

Keywords:
Digital forensics
Chat, machine learning
Sexual solicitations
Child sexual exploitation
Internet crimes against children
Linguistic inquiry and word count

ABSTRACT

Investigating crimes against children, specifically sexual solicitations, are complicated because not all offenders are contact-driven, meaning they want to meet the minor for sex in the physical world; instead, some offenders are fantasy-driven, in that they are more interested in cybersex and role-play. In addition, the sheer volume of cases involving the online sexual solicitation of minors makes it difficult for law enforcement to determine whether an offender is contact-driven vs. fantasy-driven. However, research shows that there are language-based differences between minors and contact-driven offenders vs. fantasy driven-offenders. Thus, we developed the Chat Analysis Triage Tool (CATT), a forensically sound investigative tool that, based on natural language processing methods, analyzes and compares chats between minors and contact-driven vs. non-contract driven offenders. Using an SVM classifier, we were successful in differentiating the classes based on character trigrams. In a matter of seconds, the existing algorithms provide an identification of an offender's risk level based on the likelihood of contact offending as inferred from the model, which assists law enforcement in their ability to triage and prioritize cases involving the sexual solicitation of minors.

According to the National Center for Missing and Exploited Children [1], online solicitation of minors falls into three categories: 1) sexual, request to engage in unwanted sexual activates or sexual talk; 2) aggressive, involved actual and/or attempted offline contact; and 3) distressing, youths stated they were afraid after the incident. The National Center for Missing and Exploited Children [1] examined 5863 CyberTipline reports made in 2015 regarding the online enticement of minors, which includes both sexual and aggressive solicitation. For a portion of these tips ($n = 3592$), the report examined the offenders' goals and determined that the majority (60%) wanted to obtain sexually explicit images from the minors; 32% of the offenders wanted to meet and have sexual contact with the minors while 8% were interested in sexual conversation and/or role-play [1].

In response to the increasing number of children experiencing crimes via the Internet, a nationwide network of more than 4500 law enforcement agencies was created in 1998 known as the Internet Crimes Against Children (ICAC) task force. Since its conception, the ICAC task force has investigated more than 775,000 complaints of alleged child sexual victimization, and in 2017 alone, they conducted over 66,000 investigations and 86,400 forensic exams resulting in the arrest of more than 10,300 offenders [2]. Investigating crimes against children, specifically online enticement (i.e., sexual solicitation), are complicated because internet child sex offenders are not homogenous (see Refs. [3,4,17]). As previously mentioned, their motivations for contacting children via the internet vary – some offenders are contact-driven (i.e., desire to meet for sex in the physical world); whereas, others are fantasy-driven (i.e., desire cybersex and role-play only [5]). In addition, previous research on sexual solicitation has focused on either distinguishing predatory vs non-predatory chats or a linguistic analysis of the grooming strategies used by predators as a whole [6–8]. However, law enforcement agencies are overwhelmed by the sheer volume of cases, and the data within each case, that needs to be processed. Thus, the overall goal was to assist law enforcement in their ability to triage and prioritize cases involving the sexual solicitation of minors by identifying differences between contact-driven and fantasy-driven offenders [9].

Since the grooming process involves language, Chiu et al. [10] examined whether there were language-based differences

* Corresponding author at: Purdue University, Department of Computer & Information Technology, Knoy Hall, Room 225, 401 N. Grant St., West Lafayette, IN 47907, United States.
E-mail addresses: kspellar@purdue.edu (K.C. Seigfried-Spellar), rogersmk@purdue.edu (M.K. Rogers), jtaylor1@purdue.edu (J.T. Rayz), yang798@purdue.edu (S.-F. Yang), kmisra@purdue.edu (K. Misra), tringenb@purdue.edu (T. Ringenberg).

https://doi.org/10.1016/j.forsciint.2019.02.028
0379-0738/

between minors and contact-driven vs. fantasy-driven child sex offenders. Specifically, Chiu et al. [10] hypothesized that contact-driven offenders were more likely to use self-disclosures as a grooming tactic compared to fantasy-driven offenders. Self-disclosures help to build trust in relationships when one person discloses personal experiences and emotions which thereby encourages the other individual to reciprocate (see Singer [11]). In addition, research shows that self-disclosures which involve negative experiences and emotions are more likely to build trust and intimacy in relationships (see Ref. [18]).

Using actual chats (between minors and offenders) obtained from Ventura County Sheriff's Department and Internet Crimes Against Children task forces nationwide, Chiu et al. [10] examined 36,029 words in 4353 messages within 107 anonymized online chat sessions by 21 people, specifically 12 youths and 9 offenders (5 contact-driven and 4 fantasy-driven), using linguistic inquiry and word count (LIWC) and statistical discourse analysis. Results showed that the chats between contact-driven offenders and minors were more likely to include first person pronouns, negative emotions, and positive emotions, compared to chats between minors and fantasy-driven offenders [10]. These results suggested that contact-driven offenders use self-disclosures as a grooming tactic. In addition, it was more likely that the self-disclosures from contact-driven offenders would elicit follow-up self-disclosures from their targeted youths. Chiu et al. [10] concluded that self-disclosures distinguished contact-driven from fantasy-driven offenders, and the self-disclosures occurred early enough that it was possible to detect the contact offenders before they meet-up with the minor in the physical world [10].

Based on Chiu et al.'s [10] preliminary findings, the current research team believed it was possible to develop a forensically sound investigative tool that, based on natural language processing (NLP) methods, analyzed and compared chats between minors and contact-driven vs. fantasy-driven offenders [12,13]. We extended the Chiu et al. study by using machine learning approaches on n-gram based features that captured misspellings and non-standard vocabulary, as well as use of self-disclosures. This tool is referred to as the Chat Analysis Triage Tool (CATT).

CATT uses a text classification model that was trained to predict whether a child sex offender was contact-driven or fantasy-driven. The model was trained on a dataset from *Perverted Justice*, a website containing archives of chat logs between child sex offenders and adults, who are volunteers pretending to be minors or "decoys" that carryout sting operations. 271 chat logs were manually labeled by subject matter experts to be either chats where the offender showed-up to meet the decoy in the physical world (i.e., "show"), or the offender never showed-up for a physical meeting (i.e., "no-show"). The offenders in each of the 271 chats were unique. Decoys were contacted by potential offenders after posing as minors in regional public chat rooms. When contact was made, the conversation would move to a one-on-one platform, which varied from text messaging to instant message services. Once a conversation was terminated or the offender was prosecuted, the full transcript of the conversation between the offender and the decoy was posted to the Perverted Justice website; summaries of the conversation, the case, and inline chat comments were also included. Both the entire chat available on Perverted Justice and the decoy's summary of the chat were used to determine the label of the chat.

"Show" chats included those in which the decoy explicitly stated the offender showed up to the sting location to meet the decoy. Additionally, chats were labeled as "show" if there was evidence in the chat showing the offender had left the chat to meet the decoy. Chats were labeled as "no-show" for the following cases: the decoy stated the offender never showed, the decoy indicated that the conversation was terminated early by either the decoy or the offender, or there was no evidence in the chat which showed that the offender had left to meet the decoy. An annotation schema was created and confirmed by two researchers (both are co-authors of the paper). One annotator labeled all 271 chats based on the above criteria; the second annotator verified that the labels were correct. There was no disagreement with the original labeling.

The chats were then processed into a document feature matrix where each chat was a document and the term-frequency inverse-document-frequency of each character trigram within the corpus were the features. The character trigrams were found to be optimally suited for classification as the offenders deliberately misspelled words, combined words, or used non-standard but relatively consistent vocabulary.

Next, a support vector machine [14] algorithm (SVM) with a radial kernel was trained on this document feature matrix to estimate the likelihood of a chat resulting in a "show" vs a "no show" (the positive class being a no show). To prevent overfitting due to a small dataset, we used a nested 10-fold cross validation method where the document feature matrix was split into training and testing sets randomly with a 90–10 ratio for each of the 10 outer folds. For each outer fold, the training set was further split into training and testing in order to select the best features using the information gain [15] criterion, as well as tune the cost-hyperparameter of the SVM algorithm. The inner fold models were trained and the average accuracy for each inner fold set was computed. The same models were then evaluated on the test set of the outer folds. Based on this evaluation scheme, the best model was determined as the least overfitted one, namely, where the inner fold test accuracy value and the outer fold test accuracy value were most comparable. The final model, with 87.1% average inner fold test accuracy (*Precision* = 0.819, *Recall* = 0.62, *F1* = 0.634, *MCC* = 0.613) and 86.4% average outer fold test accuracy (*Precision* = 0.975, *Recall* = 0.467, *F1* = 0.620, *MCC* = 0.623), was selected. The risk level for a contact offense is judged determined on the model's predicted probabilities for "show" vs "no-show".

Ideally, machine learning approaches would use a large amount of data to detect patterns of conversation between contact-driven and fantasy-driven offenders. The state of the art techniques that exist today are capable of comparing conversations based on deep level analyses that detect some of the meaning of the underlying text, as opposed to capturing shallow parameters, such a particular words, phrases, or syntactic constraints. However, even with limited data, it is possible to use supervised methods that differentiate between several groups of participants' speech, such as decoys (i.e., adults who pretend to be minors) vs. minors as they talk to real offenders, or contact-driven vs. fantasy-driven offenders in conversations with actual minors. The accuracy results are based on a relatively small sample and thus it is hard to give numbers that would generalize to a larger sample for various geographic locations, gender, and age; however, they are significantly above a baseline of random selection.

Finally, some offenders are aware of the capabilities of computational systems that are used to identify them, and they attempt to outwit the technology by combining words or misspelling them. This is done in such a way that a human can still understand the meaning of the text, based on – most of the time – similar pronunciation between the written text and regular words. The use of characters instead of regular words in classification ensures that Chat Analysis Triage Tool is not affected by such anomalies. In conclusion, we envision CATT assisting the law enforcement community in several ways. The existing algorithms allow an identification of a risk level, leading to a triage mechanism for prioritizing child sexual solicitation cases with a higher risk level for contact offending. When internet sex offenders contact children online, they often contact multiple children, referred to as

"spray and prey" [16]. By prioritizing those cases involving contact offenders, police officers could prevent offenders from meeting children with whom they are already speaking with, since as mentioned before, most offenders communicate with multiple victims at the same time. In addition, CATT may serve as an educational tool for under cover law enforcement investigating child sexual solicitation cases. Finally, CATT is capable of processing large amounts of data, including thousands of lines of text, in a matter of seconds – much faster than the time it would take for law enforcement to process and examine the data manually. Overall, the research team continues to work with law enforcement to obtain more real-world cases involving chats between minors and child sex offenders, as well as building additional modules in CATT that will examine other features of interest to law enforcement working internet crimes against children.

Acknowledgment

This project was funded through an internet grant, 2016 Purdue Polytechnic Institute Seed Grant program, at Purdue University.

References

[1] National Center for Missing and Exploited Children, The online enticement of children: An in-depth analysis of CyberTipline reports Retrieved from, (2017) . https://www.missingkids.org.
[2] Internet Crimes Against Children Task Force Program. Program Summary, Office of Juvenile Justice and Delinquency Prevention, n.d. Retrieved March 11, 2018 from www.ojjdp.gov.
[3] M. Henshaw, J.R.P. Ogloff, J.A. Clough, Looking beyond the screen: a critical review of the literature on the online child pornography offender, Sexual Abuse 29 (5) (2017) 416–445.
[4] K.C. Seigfried-Spellar, V. Soldino, Child Sexual Exploitation: Introduction to a Global Problem. In: T. Holt, A. Bossler (Eds.), Palgrave Handbook of International Cybercrime and Cyberdeviance. Palgrave Macmillan (in press).
[5] P. Briggs, W.T. Simon, S. Simonsen, An exploratory study of internet initiated sexual offenses and the chat room sex offender: has the internet enabled a new typology of sex offender? Sexual Abuse 23 (1) (2011) 72–91.
[6] P.J. Black, M. Wollis, M. Woodworth, J.T. Hancock, A linguistic analysis of grooming strategies of online child sex offenders: implications for our understanding of predatory sexual behavior in an increasingly computer-mediated world, Child Abuse Negl. 44 (2015) 140–149.
[7] M. Ebrahimi, C.Y. Suen, O. Ormandjieva, Detecting predatory conversations in social media by deep convolutional neural networks, Digital Invest. 18 (2016) 33–49.
[8] H.J. Escalante, E. Villatoro-Tello, S.E. Garza, A.P. López-Monroy, M. Montes-y-Gómez, L. Villaseñor-Pineda, Early detection of deception and aggressiveness using profile-based representations, Expert Syst. Appl. 89 (2017) 99–111.
[9] K.C. Seigfried-Spellar, T.R. Ringenberg, M.M. Chiu, M.K. Rogers, Distinguishing contact child sex offenders vs. non-contact solicitors: toward a digital forensics tool for automatic analysis of their chats with minors, Workshop Presented at the Association of Law Enforcement Intelligence Unit (LEIU)/International Association of Law Enforcement Intelligence Analysts (IALEIA) 2017 Annual Training Event (2017).
[10] M.M. Chiu, K.C. Seigfried-Spellar, T.R. Ringenberg, Detecting contact vs. fantasy online sexual offenders in chats with minors: statistical discourse analysis of self-disclosure and emotion words, Child Abuse Negl. 81 (2018) 128–138.
[11] A.J. Singer, Teaching to Learn, Learning To Teach, Routledge, New York, NY, 2013.
[12] Purdue Newsroom, Algorithm tool works to silence online chatroom sex predators (2018). April 17. Retrieved from https://www.purdue.edu/newsroom/releases/2018/Q2/algorithm-tool-works-to-silence-online-chatroom-sex-predators.html.
[13] K.C. Seigfried-Spellar, T.R. Ringenberg, M.K. Rogers, S.-F. Yang, K. Misra, J.M. Rayz, Digital forensic tool that analyzes offender-minor chats, Workshop Presented at the Association of Law Enforcement Intelligence Unit (LEIU)/International Association of Law Enforcement Intelligence Analysts (IALEIA) 2018 Annual Training Event (2018).
[14] C. Cortes, V. Vapnik, Support-vector networks, Mach Learn 20 (3) (1995) 273–297.
[15] J.T. Kent, Information gain and a general measure of correlation, Biometrika 70 (1) (1983) 163–173.
[16] P. de Santisteban, J. del Hoyo, M.Á. Alcázar-Córcoles, M. Gámez-Guadix, Progression, maintenance, and feedback of online child sexual grooming: a qualitative analysis of online predators, Child Abuse Negl. 80 (2018) 203–215.
[17] K.M. Babchishin, R.K. Hanson, C.A. Hermann, The characteristics of online sex offenders: a meta-analysis, Sexual Abuse 23 (1) (2011) 92–123.
[18] N.N. Bazarova, Public intimacy: disclosure interpretation and social judgments on Facebook, J. Commun. 62 (5) (2012) 815–832.