



ISSN:2229-6107



**INTERNATIONAL JOURNAL OF
PURE AND APPLIED SCIENCE & TECHNOLOGY**

E-mail :
editor.ijpast@gmail.com
editor@ijpast.in

www.ijpast.in

Using Logical Models to Categorize Personal Hygiene Discussions Online

Mrs. G. Swarna Latha, Dr. R. Rambabu , Mr. Ch.Gopi

Abstract

The goal of this study is to develop a logistic mathematical model for the purpose of locating online grooming incidents. A plethora of variables underscore the vital nature of our work, such as the explosive growth of social media, the cyberepidemic of sexual assault, and the general proliferation of cybercrime. impacts on a child's body are complicated and vary. In 2009 and 2010, the UK's Home Office's Serious Organised Crime Directorate received the most complaints about suspected online conduct, including grooming, related to child exploitation and online protection. Our analysis of more than 160 scripted online interactions allows us to pinpoint the elements of a grooming discourse.

1. Introduction

Our main goal is to create a mathematical model that can identify scripts for online conversations that involve grooming chats. According to the Cambridge Online Dictionary 1.1, "grooming discussion" is defined as "the illegal practice of befriending a child, usually online, with the purpose of encouraging the youngster to participate in sexually exploitative conduct." This artwork was inspired by a number of things. One is the remarkable rise in the number of people utilizing the Internet globally in recent years. As an example, consider the fact that 67% of American houses with children also have internet connection; for children ages 12 to 17, this percentage jumps to 84%, and for those ages 18 to 24, it reaches 97%. With the growth of social media and websites, as well Our third point is that there is an increase in cybercrime, including online grooming. Fourth, there might be psychological, behavioral, emotional, and legal repercussions for sexually abusing kids. After compiling complaints from 2009 and 2010, the Child Exploitation and Online Protection Service (CEOP)

discovered that the most often reported suspected Internet behavior was online grooming. Inappropriate sexual approaches and encouraging young people to engage in sexual behavior are examples of internet grooming. The Child Exploitation and Online Protection Centre (CEOP) was founded by the UK's Home Office's Serious Organised Crime Agency (SoCA) to research the prevalence of sexual crimes against children, both online and offline, and to provide guidance on the creation of safeguards. CEOP's observations indicate that the offenders most likely possess a thorough understanding of IT and the advantages and disadvantages of the legal system. Anyone, for good or ill, may do anything with the aid of the Internet. As a result, it could be simpler for sexual offenders to adopt false identities and avoid detection. Researching potential victims is made easier for sexual predators by online databases. Even though sexual offenders typically exploit unanticipated circumstances, they nonetheless need to put in some time and effort to arrange their crimes.

Associate Professor , Professor & HOD, Assistant Professor

Department of Computer Science & Engineering,

Rajamahendri Institute of Engineering & Technology, Rajamahendravaram.

5. Before they may attack, they must select, pursue, and seize their prey. Because they pose a continuous risk to society, over three quarters of sexual offenders need to serve out the remainder of their lives behind bars.6. Nevertheless, traces. This digital data might be valuable in forensic investigations concerning sexual offenders. Cited in the 2012 publication by Al-Zaidy and Fung. Additionally, law enforcement may use digital footprints to apprehend criminals. The data collected digitally will lend itself well to the mathematical model that the police are working to implement.

2. Approaches to the Study

This research was carried out using the following methodology. We started by visiting <http://www.perverted-justice.com/> and downloading 111 chat scripts at random, along with 48 scripts from www.literotika.com. The now-defunct site included over 500 conversations between juvenile victims or law enforcement and predators who groom children. Police disguising as underage persons. Internet grooming had already shown to be the case in every encounter. As for the second, people may talk freely about their sexual urges there without worrying about getting in trouble. A similar approach has been used by Elzinga et al.⁷ and Wollis⁸. There will be a total of 119 conversations used for testing purposes, with 100 serving as training data. First, we built the tf-idf matrix, which combines the word frequency with the inverse document frequency; second, we will randomly choose samples from both datasets to measure the prevalence of grooming qualities. As we'll see in Section 3, there are twenty separate groups that deal with different parts of one's personal appearance. One example of a scripted discussion that exemplifies the grooming character is shown in Table 1. This is where speech script no. 1 represents everything with the exception of grooming characteristic 15. $Y = 1$ because we are unquestionably talking about grooming. The script will initialize the grooming information to 0 if they are not present. Only scripts that include grooming dialogues have $Y = 0$.

Table 1: An example of the grooming conversation's characterization.

Script No	$X_k, k = 1, \dots, 20$																				Y
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
2										...											0

3. Supporting Theory

Methods for Grooming

Because of its intricacy, the grooming process has been broken down into several steps by many researchers^{4, 9–11}. O'Connell¹⁰ distinguished between the following stages. The pedophile first befriends the child by asking them many questions and getting to know them. Pedophiles vary in the amount of time they spend engaging in this behavior. Continuing to cultivate the friendship is a part of the second stage of relationship growth. Here is when the grownup may strike up a conversation about the child's day-to-day existence, including school and family. I know it's not true for all adults, but when I do, I generally stay in contact with the child to pretend we're best friends. Stage three risk assessment: this is the part of the conversation when the pedophiles ask where the computer is and how many people use it. The pedophile is testing the waters to see whether anybody—the child's parents, guardians, or older siblings—would get wind of his illicit activity. Fourth, the velocity of conversations often changes into the exclusive phase after the risk assessment phase. Pedophiles may trick their victims into thinking they understand what an adult said when they steer the conversation in a certain direction. Telling each other a tiny secret is definitely an effort to build trust at this stage. The Sexual Phase, Phase 5. Questions like "Have you ever been kissed?" or "Do you have ever touched yourself?" are common among pedophiles in this situation. An adult will often place the introduction at this stage of the dialog in such a way that there is a strong sense of shared push. It was hard for the child to notice the change in level of difficulty since they had never gone into the topic previously. The six-step process was also proposed by Welner¹¹. Abusers often follow a certain pattern: they seek out weak children, gain their trust, meet a need, isolate them, sexualize their bond, and finally, exert power. Locating vulnerable children,

enlisting the help of their classmates, acclimating them to physical contact, placing them in an isolated area, and instilling a sense of responsibility are all components of the Lanning⁴ grooming approach. Table 2 shows the stages that Gupta⁹ suggests, along with a short explanation of what they are.

Table 2. Descriptor for Stage of Online Grooming⁹.

	Stage					
	Friendship	Relationship	Risk Assessment	Exclusivity	Sexual	Conclusion
Descriptor 1	Exchange email / picture / web-cam information (early in chat)	Exchange email address / picture / web-cam information (later in chat)	Checking child's parents are around or who all else use computer	Feeling of love and exclusiveness expressed	Giving body and figure description	Arranging for a day, date, time and location to meet in person
Descriptor 2	Talking about boy friend / girl (early in chat)	Giving soft compliments, e.g., sweetie, cutie	Asking the child to delete their chat logs, ensuring nobody else had password of child's account	Describing sexual activity and experiences to the child	Becoming boys friend / girlfriend of each others	Discussing about the commute to the meeting point
Descriptor 3	Giving information about other accounts and online profiles of the child	Talking about a child's hobbies, activities and interest	Checking if child is fine with seeing an older man/woman	Giving strong compliments e.g., you are sweetheart	Exchanging picture of sexual nature or body parts	Ensuring child will come alone to meet
Descriptor 4	Asking the age / gender / location / name / personal information / details about family	School / Grade / Homework / Cell-phone number	Directly confronting to ensure that child not a cop / police agent	Built trust with the child	Giving sexual oriented compliment e.g., sexy	Deciding on what to do when they meet in person

Attributes of Personal Grooming

X1: Examining the potential dangers of a discussion
Predators usually consider the possibility that the victim's parents may overhear their conversation with a possible victim. The parents could bring up the victim's identity in discussion if they are aware of it. Run the risk of facing legal consequences because their parents reported them to the police. Potential victims are often questioned by predators regarding the computer, including who uses it, where it is stored, and whether or not the victim's parents are aware of the password for the chat application. The offender then goes on to warn the victim that there might be legal repercussions for what they did. (X2). Criminals will tell victims this so they won't be caught themselves. This is a typical strategy used by predators who wish to continue conversing with possible victims.

The presence or absence of an adult partner is another related aspect. Number four. The predator may feel safe from legal consequences if no one confides in them or reports the talk to their parents.¹²⁻¹³. Communication via channels other than the Internet (X5): Predators are now exploring options outside of online communication; they find text-based, media-based, and voice-based forms of contact more intriguing and would want to feel gratified by them. Here, predators are trying to win over their potential victims by showing them they care. They

will have a much easier time moving ahead with their victims if they are able to do this. The use of words pertaining to emotions in speech; this encompasses both potential victims and predators. Biology, anatomy, and sex-related terminology (X8): These words will be used in any discussion that aims to become sexual. The use of terminology for sexual organs that are prevalent among children is considered sexually predatory language (X9). 8.13. Sexual slang or daily language terms pertaining to private areas of the body (X10): terms used in both formal and informal contexts.⁷. Reframing (X11): Now we'll subtly include sexual elements involving predators¹⁴ and redefining sexual behavior as something other than sexual, including associating it with playing, training, or instructing¹³. Images having a sexual theme may be requested by predators from potential victims who they find attractive (X12). Imaginative or used as a weapon of blackmail, sexually explicit images may have both positive and negative effects.^{12, 13, 7, 10, and 12}. Phase X13 involves the predator communicating with the victim in an attempt to desensitize them to discussing sexual matters. For example, some predators would alter the letter "p" in "pick" to a "d" and say it was a mistake, while others would intentionally replace non-sexual phrases with sexual ones. The sexual predator will inquire about the victim's sexual desires and experiences (X14) in an effort to gauge the victim's receptivity to future sexual touches. This is also the time when the predator may ask about the victim's sexual history. Tell me how frequently you have sex, if ever. Since discussing sexuality is no longer socially unacceptable, predators assume that their sexually virgin prey will find it easier to engage in sexual behavior with them.¹². Prior to the development of predators' imaginations to the kern¹⁰ level, the first mention of sexual context was made (X15). We shall introduce the subject of fantasy, although in a very generalized way, at this stage of the conversation (X16). Fun with playing out dreams (X17): So far, so good, but no signs of intimacy have emerged in the conversation.

The Inverse Document Frequency and Term Frequency

To find out how often each phrase occurs in the discourse, we will next use a typical approach for analyzing papers called term frequency-inverse document frequency (TF-IDF). The source is Russell (2014). The main goal of this process is to find out how often certain phrases occur in a specific text or collection of texts. The goal is to get the inverse-document-frequency (IDF) of the phrase by counting the number of documents that include it.

Logistic Regression, Binary Option

The binary logistic model's generalized equation is $P = \frac{1}{1 + e^{-X\beta}}$. The variables $X_1, X_2, X_3, X_4, X_5, X_k, X$, and Y may all be either zero or one, making them binary variables. There are k -type aspects of grooming chats ($k=20$) if $X_k = 1$. For the model to interpret it as a grooming conversation, Y must be equal to 1.

Metrics for Performance

Table 3 shows the confusion matrix that is used to evaluate the developed logistic model. Also taken into account is the model's validity, as stated in Eq. (1). (2).

The meaning of "contingency table" is listed in Table 3.

		Actual	
		Yes	No
Prediction	Yes	True Positive (TP)	False Positive (FP)
	No	False Negative (FN)	True Negative (TN)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (2)$$

4. Results

Feature Extractions

After identifying the grooming aspects of every conversation screenplay, they were grouped into 20 distinct groups (see to the table below for details). Table 4 summarizes the grooming characteristics identified in one hundred conversation transcripts. Some interesting conclusions derived from the data set are as follows. Among the most common grooming practices are the following: "utilizing words in biology, body, and sexual category," "introducing sexual stage," "utilizing words in feeling category," "arranging further contact and meeting," "telling the sexual preference or desire and sexual experience," and "calling intimate part using popular name or slang word," in that order. Discussions on personal hygiene almost seldom use parenting-related vocabulary. Of the twenty-three discussions that do not touch on grooming, seven do not have access to any of the attributes related to it. Nonetheless, a great deal of grooming signals are detected in otherwise non-specific speech. As a result, classifying discussions on appropriate grooming procedures may be challenging. Words that

aren't typically associated with grooming, such as "biology, body, and sexual," "introduced sexual stage," "fantasy enactment initial stage," and "fantasy enactment-based activity," are frequently used in conversations about grooming.

Development of Models

Before the logistic model is set up, all of the independent variables are checked for potential associations with the dependent variable Y . You may measure the strength of a link using statistical tests like the paired t-test, Spearman's rank correlation coefficient, and Pearson's correlation coefficient. However, the first choice is selected since all the data is binary. The paired t-test results for all independent variables are shown in Table 5. The p-values span a wide range, from 0.000 to 0.9999. The significance of the p-value suggests that the independent variables of interest are those that significantly affect the dependent variable. A p-value for an independent variable is considered statistically significant if it is less than a preset threshold; for this example, we will use a p-value of 0.25. According to this criteria, the logistic model development procedure does not need the following components: In terms of the X-series, 5, 11, 12, 14, 15,

It is possible to simplify Eq. (1) using this data.

$$\ln\left(\frac{P}{1-P}\right) = b_0 + b_5X_5 + b_{11}X_{11} + b_{12}X_{12} + b_{14}X_{14} + b_{15}X_{15} \quad (3)$$

To get the model coefficients $b_0, b_5, b_{11}, b_{12}, b_{14},$ and b_{15} , step-wise regression uses three methods: the Enter Method, the Forward Stepwise Method, and the Backward Stepwise Method. You can see the results of the sequential regression analysis in Table 6. The calculated model coefficients ($b_0, b_5, b_{11}, b_{12}, b_{14},$ and b_{15}), together with the techniques and results, are listed in the table. With focus on the most relevant variable in each model and its degree of relevance. The most crucial component is undeniably the one that stands out with the highest p-value and the lowest coefficient value in comparison to the other model variables. An intuitive assessment is necessary for the paired t-test result of Eq. (3). Consequently, it follows that an accurate independent variable for identifying grooming conversations is necessary. **Table 6. Logistic models have been developed for categorizing online grooming discourse.**

Model No.	b ₅	b ₁₁	b ₁₂	b ₁₄	b ₁₅	b ₀	Critical Variable	p-value
1	3.514	1.884	1.201	3.409	3.654	-5.742	X ₁₂	0.315
2	3.784	-	-	4.211	3.699	-5.570	X ₁₅	0.083
3	3.231	-	-	4.313	-	-2.026	X ₅	0.001
4	3.231	-	-	4.313	-	-2.026	X ₅	0.001

X11 and X12 as our means. These conclusions were possible because of the computed p-value and Wald statistic.

Table 7. X5, X11, X12, X14, and X15 are five independent variables that are statistically reviewed.

Variable	B	S.E.	Wald	df	p-value	Exp(B)
X ₅	3.514	1.198	8.600	1	0.003	33.583
X ₁₁	1.884	1.338	1.981	1	0.159	6.578
X ₁₂	1.201	1.195	1.010	1	0.315	3.322
X ₁₄	3.409	0.968	12.392	1	0.000	30.228
X ₁₅	3.654	1.914	3.646	1	0.056	38.648
Constant	-5.742	2.068	7.708	1	0.005	0.003

The results of the statistical tests for X5, X14, and X15 are summarized in Table 8.

Variable	B	S.E.	Wald	df	p-value	Exp(B)
X ₅	3.784	1.170	10.453	1	0.001	43.986
X ₁₄	4.211	0.904	21.690	1	0.000	67.416
X ₁₅	3.699	2.133	3.008	1	0.083	40.417
Constant	-5.57	2.240	6.182	1	0.013	0.004

Table 9 presents a summary of the statistical tests conducted on X5 and X14, two independent variables.

Variable	B	S.E.	Wald	df	p-value	Exp(B)
X ₅	3.232	0.935	11.935	1	0.001	25.317
X ₁₄	4.313	0.890	23.463	1	0.000	74.670
Constant	-2.026	0.533	14.472	1	0.000	0.1320

Evaluation of the Model

The accuracy and recall (Table 3) are used to evaluate the effectiveness of the Table 6 model. Table 10 shows the outcomes of 100 scripted discussions from the training set, whereas Table 11 shows the outcomes from 59 scripted interactions from the testing set. Keep in mind that while evaluating performance, only the third model is considered. While the first and second models may improve recollection, accuracy is likely to suffer. There is a 92% success rate on the training set and a 95% success rate on the test set. So, it seems like the third model is pretty spot on.

Training set results for the logistic model are shown in Table 10.

When discussing personal hygiene, it should be used more often than in other contexts. According to Table 4, X14 appears in 85% of scripts pertaining to grooming talks but in a mere 4% of scripts pertaining to all other topics. Although X5 is not discussed at all outside of conversations regarding grooming (94%), it does appear 67% of the time when discussing grooming. The variable X11 is mentioned in 60% of grooming-related interactions and not in 97%. Two intriguing considerations are X12 and X15. While 94% of all discussions do not include variable X12, only 34% of grooming interactions do. X15, on the other hand, is brought up 97% of the time while discussing personal hygiene, but just 18% of the time when discussing other topics. Based on these results, it seems that X15 is a good predictor of speech that pertains to grooming, whereas X12 is more accurate for general discourse.

Five models are shown in Table 6, each with its unique set of independent variable coefficients: b₅ = 3.514, b₁₁ = 1.884, b₁₂ = 1.201, b₁₄ = 3.409, and b₁₅ = 3.654. The associated variable's effect on the model's prediction is precisely described by the numerical value. The values of the two smallest coefficients, b₁₁ and b₁₂, deviate significantly from the values of the other coefficients. X12 is the least significant variable with a coefficient of 1.202 and a p-value of 0.315. The second one, with less weight, is variable X11. X11 and X12 are absent from the second model. In this model, the coefficients are: X15, the least significant variable in the model, has a p-value of 0.083, whereas b₅ = 3.784, b₁₄ = 4.211, and b₁₅ = 3.699. The elimination of X11 and X12 reduces the model uncertainty from 31.5% to 8.3%. Further reduction of the number of independent variables produces a model with an extraordinarily low p-value. With only two variables (X5 and X14) and a small margin of error (0.1%), the model is constructed.

Important Findings from Models 1, 2, and 3's Tests We will go over the results of the significance tests for Models 1, 2, and 3 further down. The results of the tests are shown in Tables 7, 8, and 9, in that sequence. The data suggests that X5 and X14 are the two most crucial variables in all of the analyses. X15 is the second most significant variable, and after determining its importance, we use

		Actual		Total
		Predator	Non predator	
Prediction	Predator	63 (94%)	4 (6%)	67
	Non predator	4 (12%)	29 (88%)	33

Table 11. Performance of the logistic model on the testing set

		Actual		Total
		Predator	Non predator	
Prediction	Predator	43 (96%)	2 (4%)	45
	Non predator	1 (7%)	13 (93%)	14

5. Conclusions

The objective of this research is to create a logistical mathematical model that can assess whether a scripted online encounter is grooming. This effort is vital since crime has increased in tandem with the advent of social media. Five major points of the grooming conversation have been identified by the study. Questions such as "What's another method to get in touch?," "What about reframing?" and "Are you interested in seeing some hot pictures?" as well as "Explaining your sexual preferences, desires, and experiences?" and "Introducing the sexual stage?" are examples of these characteristics. And using these numbers, we were able to build a mathematical model of logistics. Based on an examination of the model's performance using the training data set of 100 scripts and the testing data set of 59 scripts, the model can identify grooming conversations with 95% accuracy, with 96% true positive and 93% true negative. There were only 4% false positives and 7% false negatives.

Works Cited

1. The second picture is available at <http://dictionary.cambridge.org/>, the website of the Institute of Health Economics' Cambridge Online Dictionary. An overview of the literature on the topic of online child sex abuse. 2010 was the third TK's cruise. Alberta is a province in Canada. Sexual abuse of children. Third Edition of Helping Children at Home and in the Classroom: Resources for Parents and Teachers from Participating Psychologists, 2010.

4. KV Lanning gives specialists looking into child sexual exploitation a presentation titled Child Molesters: A Behavioral Study. Under the direction of the US Department of Justice's Office of Justice

Programmes, which is in charge of juvenile justice and penalty avoidance. The Psychology of Criminal Behavior by Andrews and Bonta (2010). Matthew Bender & Company, Inc. released this book in 2010 in New Providence, NJ. Sixth, M.

Psychcentral.com/lib/definition/sexual-addiction/000748. The return date was August 6, 2014.

7. Elzinga, Wolff, and Poelmans, J.: Using Temporal Relational Semantic Systems to Analyze Paedophile Chat Conversations. The proceedings of the 2012 European Intelligence and Security Informatics Conference (EISIC) are located on pages 242-249. Using a linguistic perspective, Follis, M. A. (2008) assessed online predator grooming. Taken from <http://dspace.library.cornell.edu/bitstream/1813/23125/2/Wollis%20Melissa%20Research%20Honors%20Thesis.pdf> by the University of Cornell, Department of Social Sciences, Research Report 2011, 2011, College of Agriculture and Life Sciences Division.

Online grooming is defined by research by Gupta, Kumaraguru, and Ashish (2018) as a trait that distinguishes internet peer discussions. Computer Science Archive. The preprint was made available in 2012 with the following DOI: 1208.4324. 10. O'Connel, R. Report on Scientific Findings: A categorization system for cybersex exploitation and cybergrooming. Release of