ICIC Express Letters ICIC International ©2019 ISSN 1881-803X
Volume **13**, Number **12**, December **2019** pp. 1087–1095

ANALYTIC HIERARCHY PROCESS IN DETECTING PROBABILITY OF ONLINE CHILD GROOMING USING NEURAL NETWORK

Derwin Suhartono[1], Wikaria Gazali[2] and Keith Ferdinand Mamahit[2]

[1]Computer Science Department
[2]Mathematics Department
School of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ dsuhartono; wikaria }@binus.edu; keithferdinandm@gmail.com

Received May 2019; accepted August 2019

Abstract. *The goal of this research is to detect probability of online child grooming by generating percentage score from online conversation. We applied 2 methods, one for classifying the conversation type and the other for calculating the online grooming percentage. We used neural network to classify the conversation into several classes. As for calculating the percentage, we used analytic hierarchy process (AHP) to find grooming score based on priority score for each class. This research focused more on online conversation that consists of two people and used English as the main language. The results can be used to determine whether the conversation is considered as groom or not.*

Keywords: Online child grooming, Neural network, Analytic hierarchy process, Online conversation, Priority score

1. **Introduction.** Most people are very familiar with the fact that technology is growing rapidly. In the past, we must wait for days to weeks only to send one message, yet now we can send many messages just in seconds. This technology movement makes life much easier. Along with positive impact of technology, we do have negative side as well; it is criminal. Society is sometimes careless to the issue that technology development drives to the growth in the number of criminal acts. One of them that is still currently happening is online child grooming.

Online child grooming is an act of crime to groom children with the purpose of sexual abuse or human trafficking. Grooming means an attempt made by pedophiles such that they can get closer to the victims by keeping communicating with them. It is intended in order that victims can be comfortable with. There are not so many handling methods to identify and to prevent the growth rate of this crime. This happened because people do not really understand how this crime occurs [1]. Online child grooming does not happen instantly. Pedophiles need to take days and even longer time to make their victim be interested and attached with them. Therefore, this process is very possible to be identified earlier. We can prevent it from happening if we do not act nonchalantly with the situation around, especially within our inner circle.

In this research, we developed an application which can be used to detect the percentage score of online child grooming. Some research questions that have been formulated are (1) what the indicators of online child grooming in such conversation are, and (2) how the proposed method can solve the problems. The objectives of this research are to analyze conversation which contains online child grooming and to classify the conversations into several categories. We used neural network architecture to classify sentences within chatting platform into some classes. Subsequently, we used the analytic hierarchy process

DOI: 10.24507/icicel.13.12.1087

(AHP) [2] to calculate grooming percentage score from the chats. AHP is a theory and methodology for relative measurement. In relative measurement, the interest is not in the exact measurement of some quantities, but rather on the proportions between them [3]. It has been used as well for determining barriers in implementing material efficiency strategy [4].

Result of this research will be presented in form of percentage score which is a measurement of how far a pedophile has done certain stages of actions. The calculation will use AHP to calculate the priority value from each stage.

2. **Related Research.** To the best of our knowledge, similar studies have been done in other countries except Indonesia. K-nearest neighbors (KNN) classifier was used to find out which conversation was done by pedophile and which one was not [5]. Pendar divided the conversation into 2 (two) types: conversation by pedophile and conversation by victim. There was no correction done to the conversation, except for words with repetition letters, such as "*noooo*" or "*coool*". However, this correction was not good enough, since words like "*good*" and "*sweet*" would be transformed into "*god*" or "*swet*". The result showed that 6.4% conversations by victims were detected as grooming while 7.1% conversations by pedophiles were not detected as grooming. It also described that using trigram to classify the sentence could give the best result.

Similar research was conducted by developing an application called Chatcoder [6]. It is a bit different from what Pendar did. By finding unique phrases from conversation, Chatcoder attempted to find out whether a conversation consists of grooming class or not. Its library had a total of 475 unique phrases. It was divided into 8 classes based on luring communication (LCT) [7]. The experiment consisted of 2 datasets which are divided into 2 (two) conversations. The first one was a conversation between pedophile and victim; the other one was a conversation between 2 adults. Chatcoder used J48 classifier to predict whether the conversation was done by pedophile or victim. Accuracy of this experiment is 60%. It is said that the number of training data was still not enough, and they needed a better training process and a better classifier.

Experiment utilizing n-gram with some high-level features was done as well. Several examples of high-level features were emotion, neuroticism level, and lexical feature. Some other low-level features were also involved [8]. Their objective was to separate conversation between grooming conversation and adult conversation by using Naive Bayes classifier. WordNet-Affect was used to detect the emotion from conversation. The emotion consisted of Joy, Sadness, Anger, Surprise, Disgust, and Fear. The result showed that no low-level features can exceed the performance of high-level features. The accuracy was around 90%-94%. Among all high-level features, emotion was considered as the best feature.

In our research, we attempt to detect the probability of conversation containing groom. We only focus on lexical feature to detect the conversation using 8 grooming classes based on LCT. We used neural network as classifier and AHP to calculate the percentage of grooming in a conversation.

3. Proposed Method.

3.1. **Dataset.** We gathered our data for training and testing from www.perverted-justice.com (PJ), where grooming conversations are being stored. The conversations were proven as grooming conversations between 2 people. We collected 100 data: 50 data were used for training and the other 50 were used to test the program. The actors in this dataset consisted of pedophile and pseudo-victim. It was called as pseudo-victim because all victims in the conversations were not real kids. Pseudo-victim was a group of people who cooperated with police and became volunteers by disguising as kids to arrest pedophiles through online conversation.

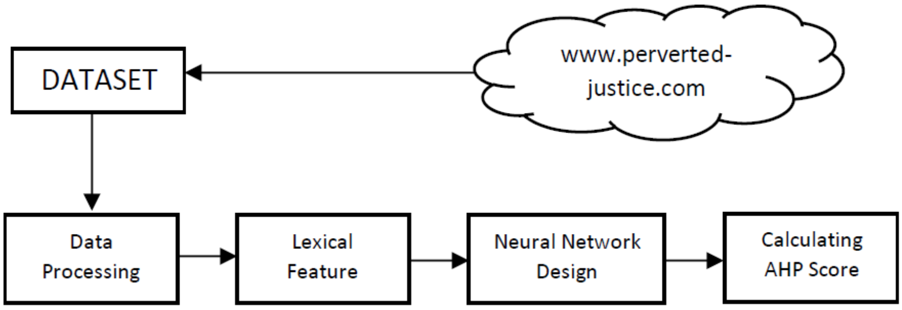
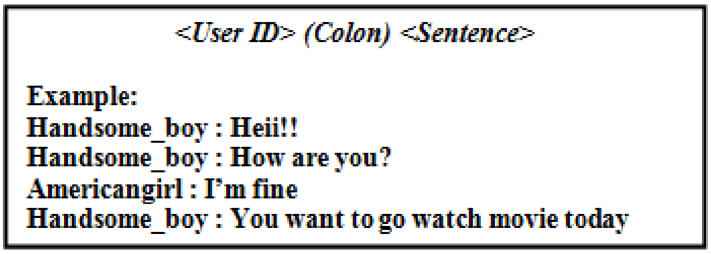


Figure 1. Flowchart of the proposed methodology

3.2. **Data preprocessing.** At first, we needed to make sure that format of the conversation can be read by the program. Every line in conversation should be formatted into "UserID: Sentence". Figure 2 is an example of the formatted conversation.



<User ID> (Colon) <Sentence>

Example:
Handsome_boy : Heii!!
Handsome_boy : How are you?
Americangirl : I'm fine
Handsome_boy : You want to go watch movie today

Figure 2. Example of conversation between 2 people

Normalizing the conversation is quite difficult since online conversations usually have their own characteristics of language. Online conversations do not stick with the correct vocabulary, nor grammar [9]. Many mistyped words are made intentionally, such that they cannot be found in dictionary. In online conversations, many emoticons are often used as well.

We separated all data into 2 conversations: conversation by pedophile and conversation by pseudo-victim. Tokenization was done afterwards. We used tokenization mechanism suggested from previous studies [10,11] such as:

- Remove non-text words, including punctuation, emoticon, and number
- Transform all words into lowercase
- Remove stop words
- Stemming and lemmatizing

This research only focused on finding the percentage score. We did not make any corrections for mistyped words or repetition letters. We did not apply any deep learning model as well.

3.3. **Lexical feature.** As we mentioned before, we used LCT to classify conversation into classes. There are 9 classes proposed by LCT, but other research said that kids nowadays just entered random chatroom and started the conversation [12]. It became the reason why we removed gaining access class. The other classes were as follows:

- Personal Information (PI)
- Relationship Information (RI)
- Activities Information (AI)
- Compliments (C)
- Communicative Desensitization (CD)
- Reframing (R)

- Isolation (I)
- Approach (A)

To classify the conversation into some classes, we used several keywords related to each class. We gathered 186 keywords for 8 classes. Table 1 shows some keywords which are stored in our library.

TABLE 1. Example of keywords stored in our library for each class

Grooming class	Sentence	Keywords
1) Personal Information (PI)	– Where do you live? – Can you send me your picture? – When is your birthday?	'Asl' (age/sex/location), 'picture', 'birthday', 'name', 'email'
2) Relationship Information (RI)	– How did you meet your boyfriend? – Did you like your boyfriend? – Have you ever date?	'boyfriend', 'girlfriend', 'date', 'breakup', 'like'
3) Activities Information (AI)	– What is your favourite color? – Did you like that book? – Where do you like to go to hangout?	'favourite', 'movie', 'book', 'hobby', 'fun', 'sparetime'
4) Compliments (C)	– You are pretty. – Nice picture. – You look good in that dress.	'pretty', 'cute', 'handsome', 'good', 'nice', 'smart'
5) Communicative Desensitization (CD)	– You have a hot body. – I would like to see you naked. – You are very welcum.	'hot', 'body', 'sex', 'naked', 'welcum', 'penis', 'fuck'
6) Reframing (R)	– I could teach you how to do it. – Do you want to play with it? – You can learn it with me.	'practice', 'teach', 'play', 'learn', 'mess around', 'help'
7) Isolation (I)	– You must be lonely? – When is your dad leave? – You have a brother or sister?	'Mom', 'lonely', 'brother', 'alone', 'leave'
8) Approach (A)	– I will come over. – How about we go to hotel? – I want to see you in real.	'meet', 'come', 'see', 'hotel', 'car', 'house'

3.4. **Neural network design.** We used backpropagation model for our neural network. For activation function, we used sigmoid unipolar. Number of nodes in input layer depends on how many words are found in 1 (one) sentence for every conversation. The number of nodes in output layer is 8 (according to number of classes). For hidden layer, the number of nodes is 17, and it comes from try and error mechanism.

For training process, we generate the weight randomly. For number of epochs, we used 40,000 epochs, and it comes from analyzing several experiments. We applied bias to updating the weight faster. Dropout was also used to find the most efficient neural network system and reduced the possibility of overfitting [13]. After all settings had been set, the training process was ready to execute. We saved the result of training by using JSON.

Subsequently, we can use the training data to classify conversation into classes. First, we need to tokenize every sentence in the conversation. After that, we need to find the value for every word or token. It will be 1 (one) if it is found in library, and 0 (zero) if it is not. Figure 4 explains us how to give value to token.

3.5. **Calculating AHP.** After getting value for every token, we can calculate the percentage score based on how many classes are found in every conversation. To do that, we need to find the score for every class first, since we believe that every class has different

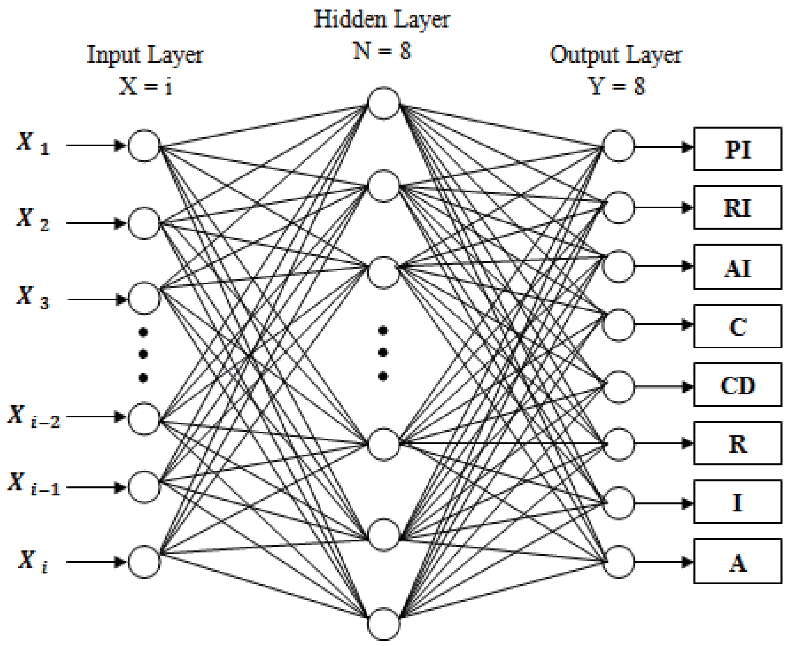


FIGURE 3. Neural network design

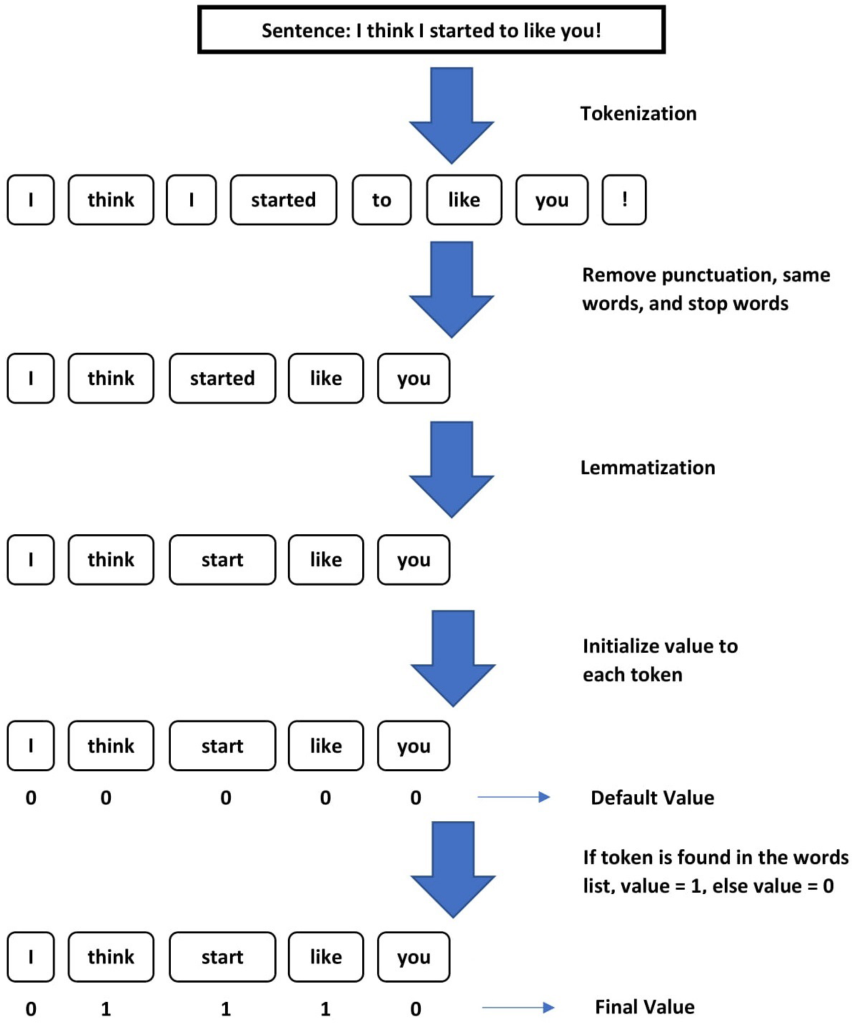

Sentence: I think I started to like you!
Tokenization
I
think
I
started
to
like
you
!
Remove punctuation, same words, and stop words
I
think
started
like
you
Lemmatization
I
think
start
like
you
Initialize value to each token
I
think
start
like
you
0 0 0 0 0
Default Value
If token is found in the words list, value = 1, else value = 0
I
think
start
like
you
0 1 1 1 0
Final Value

FIGURE 4. Step by step of how token can get value

scores. There are classes with high priority, and classes with less priority. To find them, we used AHP.

We got the priority score from the average difference between grooming data and non-grooming data for every class. Equation (1) explains how we can find the average.

$$\overline{x\%} = \left(\frac{\text{Number of lines for 1 class}}{\text{Number of lines from conversation}}\right) \times 100\% \tag{1}$$

The calculation result can be seen in Table 2. We gathered and calculated the average from 50 grooming data and 50 non-grooming data.

Table 2. Results of the difference between grooming data and non-grooming data

$\overline{x\%}$	PI	RI	AI	C	CD	R	I	A
Grooming data	2.84	2.61	1.14	4.42	5.16	1.45	3.86	6.67
Non-grooming data	2.34	2.27	1.09	3.18	1.53	1.04	3.76	5.96
Difference	0.49	0.33	0.04	1.24	3.63	0.40	0.10	0.71

We can see clearly that CD class has the highest difference which made it into our highest priority. Table 3 shows the ranking of priority descending from the top priority to least priority.

Table 3. List of priority ordered by difference

Rank	Class	Difference
1	CD	3.63038
2	C	1.24593
3	A	0.71068
4	PI	0.49459
5	R	0.40271
6	RI	0.33296
7	I	0.10054
8	AI	0.04923

Furthermore, we made pairwise comparison matrix based on differences between classes. Pairwise comparison is a process comparing 2 activities to determine which one is preferable. For example, in term of choosing fruits, banana is preferable compared to apple. Thus, 5 values will be assigned to explain the level of banana rather than to apple.

The value we gave might be a little bit subjective, but it was still acceptable considering the difference between classes. After we have set the matrix, we continued to find the priority vector. This priority vector is used to set the maximum percentage score of every class and to calculate the final score. Figure 5 and Table 4 respectively show the result of our pairwise matrix comparison and the priority vector.

The priority vectors as the output from AHP are used as the benchmark for each class. Every class has a different maximum value depending on the priority vector values. Furthermore, a value for 1% increment to result is calculated by considering how many percentages is needed from each class. Thus, average percentage value from each class is divided by priority vector value from related class. The value is taken from grooming data.

The next step is to divide the average of grooming data with the priority vector. It is to find the increment of every class used in the result. We used grooming data instead

$$\begin{array}{c|cccccccc|}
 & \textbf{PI} & \textbf{RI} & \textbf{AI} & \textbf{C} & \textbf{CD} & \textbf{R} & \textbf{I} & \textbf{A} \\
PI & 1 & 3 & 7 & 1/3 & 1/7.5 & 2 & 6 & 1/3 \\
RI & 1/3 & 1 & 6 & 1/5 & 1/7 & 1/3 & 4 & 1/5 \\
AI & 1/7 & 1/6 & 1 & 1/8 & 1/9 & 1/7 & 1/3 & 1/7.5 \\
C & 3 & 5 & 8 & 1 & 1/5 & 4 & 7 & 2 \\
CD & 7.5 & 7 & 9 & 5 & 1 & 7.5 & 8 & 6 \\
R & 1/2 & 3 & 7 & 1/4 & 1/7.5 & 1 & 5 & 1/4 \\
I & 1/6 & 1/4 & 3 & 1/7 & 1/8 & 1/53 & 1 & 1/6.5 \\
A & 3 & 5 & 7.5 & 1/2 & 1/6 & 4 & 6.5 & 1
\end{array}$$

Figure 5. Result of pairwise comparison matrix

Table 4. Result of priority vector as well as maximum percentage score of every class for magnetic properties

Class	Priority vector	Max percentage score/class
PI	0.089961475	8.9961475%
RI	0.051356081	5.1356081%
AI	0.018296972	1.8296972%
C	0.169468311	16.9468311%
CD	0.398720744	39.8720744%
R	0.073862729	7.3862729%
I	0.029257867	2.9257867%
A	0.169075821	16.9075821%

Table 5. Increment of every class

Class	Average of grooming data	Percentage of priority vector	Increment
PI	2.84319	8.9961475%	0.316045%
RI	2.61027	5.1356081%	0.508269%
AI	1.14712	1.8296972%	0.626945%
C	4.42787	16.9468311%	0.26128%
CD	5.16061	39.8720744%	0.129429%
R	1.45064	7.3862729%	0.196397%
I	3.86321	2.9257867%	1.3204%
A	6.67391	16.9075821%	0.394729%

of non-grooming or combination of both because we aimed to calculate percentage of grooming score. The results can be seen in Table 5.

For every line that is classified as ones class, the final percentage will be increased according to its class' increment. However, if the total percentage of ones class exceeds that class' priority vector, the total percentage of that class will be set to be the same as the priority vector. The final percentage can be obtained by summing up the percentage of all classes. Finally, we set the grooming threshold to 65%. If the score is less than 65%, it will not be considered as grooming. This threshold is chosen from several experiments conducted by us.

4. **Results and Discussion.** We did a test with our program to make sure whether it worked correctly or not. We used 100 data consisting of 50 grooming conversations and 50 non-grooming conversations. The result showed that from 50 grooming conversations, 16 conversations were not detected as grooming. As for non-grooming conversations, 11 conversations were detected as grooming. The accuracies of grooming conversation and non-grooming conversation were 68% and 78% respectively.

We are confident that we can increase the accuracy if we do correction to the sentence. This includes mistyped words, repetition letters, and grammar correction. The number of data for training is also not enough, since we only use 186 words as the keywords.

Overall, the program worked as we expected, even though the result was slightly worse than what we expected. The program can find out the score of grooming percentage from the conversation.

5. **Conclusions and Future Works.** From all the discussions above, we arrived at conclusions that:

- The grooming class proposed by LCT can be used to classify conversation into some classes, which later can be used to find the percentage grooming score;
- AHP can be used to find percentage score of classification data;
- Our training data for neural network which consists of 186 words are still not enough to get a good accuracy.

In the future, we will increase the capability of our program by increasing the number of training data. Also, we will do a correction with words and sentences. We will try to apply deep learning, so the program can perform better. It would be good if we can make similar program which can be used to detect Bahasa, since all conversations used in our program are still in English. Last, but not least, we take into consideration what has been done before [8]. We will try to detect the conversation based on emotions that can be found through the conversation, since the result of their research said that emotion has the best result compared to other features.

REFERENCES

[1] T. Randhawa and S. Jacobs, *Child Grooming: 'Offending All the Way through from the Start' – Exploring the Call from Law Reform*, Australia: Child Wise, 2013.

[2] T. L. Saaty, Decision making with the analytic hierarchy process, *Int. J. Services Sciences*, vol.1, no.1, pp.83-98, 2008.

[3] M. Brunelli, *Introduction to the Analytic Hierarchy Process (Springer Briefs in Operations Research)*, Springer, 2015.

[4] F. H. Ho, S. H. Abdul-Rashid and R. A. R. Ghazilla, Analytic hierarchy process-based analysis to determine the barriers to implementing a material efficiency strategy: Electrical and electronics' companies in the Malaysian context, *Sustainability*, doi:10.3390/su8101035, vol.8, no.10, 2016.

[5] N. Pendar, Towards spotting the pedophile telling victim from predator in text chats, *Proc. of the International Conference on Semantic Computing*, pp.235-241, 2007.

[6] A. Kontostathis, L. Edwards and A. Leatherman, Chatcoder: Towards the tracking and categorization of Internet predators, *Proc. of Text Mining Workshop 2009 Held in Conjunction with the 9th Siam International Conference on Data Mining (SDM 2009)*, 2009.

[7] L. N. Olson, J. L. Dagss, B. L. Ellevold and T. K. K. Rogers, Entrapping the innocent: Toward a theory of child sexual predators' luring communication, *Communication Theory*, vol.17, no.3, pp.231-251, 2007.

[8] D. Bogdanova, P. Rosso and T. Solorio, On the impact of sentiment and emotion based features in detecting online sexual predators, *Proc. of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, Korea, pp.110-118, 2012.

[9] A. Gupta, P. Kumaraguru and A. Sureka, Characterizing pedophile conversations on the Internet using online grooming, *arXiv*: 1208.4324, 2012.

[10] A. E. Cano, M. Fernandez and H. Alani, Detecting child grooming behaviour patterns on social media, *Lecture Notes in Computer Science*, vol.8851, Springer, Cham, 2014.

[11] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistic, and Speech Recognition*, 2nd Edition, Prentice Hall, 2008.

[12] I. McGhee, J. Bayzick, A. Kontostathis, L. Edwards, A. McBride and E. Jakubowski, Learning to identify Internet sexual predation, *International Journal of Electronic Commerce*, vol.15, no.3, pp.103-122, 2011.

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, vol.15, pp.1929-1958, 2014.