Modelling Fixated Discourse in Chats with Cyberpedophiles

Dasha Bogdanova
University of
Saint Petersburg
dasha.bogdanova
@gmail.com

Paolo Rosso
NLE Lab. - ELiRF,
Univ. Politécnica de Valencia
prosso@dsic.upv.es

Thamar Solorio
University of
Alabama at Birmingham
solorio@cis.uab.edu

Abstract

The ability to detect deceptive statements in predatory communications can help in the identification of sexual predators, a type of deception that is recently attracting the attention of the research community. Due to the intention of a pedophile of hiding his/her true identity (name, age, gender and location) its detection is a challenge. According to previous research, fixated discourse is one of the main characteristics inherent to the language of online sexual predation. In this paper we approach this problem by computing sex-related lexical chains spanning over the conversation. Our study shows a considerable variation in the length of sex-related lexical chains according to the nature of the corpus, which supports our belief that this could be a valuable feature in an automated pedophile detection system.

1 Introduction

Child sexual abuse is not a rare problem. The statistical analysis by the National Incident-Based Reporting System data (FBI, 1995) revealed that in the majority of all sexual assaults (67%) the victims were underage (Snyder, 2000). Child sexual abuse and pedophilia are related to each other and both are of great social concern. On the one hand, law enforcement is working on prosecuting and preventing child sexual abuse. On the other hand, psychologists and mental specialists are investigating the phenomenon of pedophilia. Even though pedophilia has been studied from different research perspectives, it remains to be a very important problem that requires further research.

The widespread availability of the Internet, and the anonymity enabled by it has brought about new forms of crime. According to the research conducted by Mitchell (2001), 19% of children have been sexually approached over the Internet. However, only 10% of such cases were reported to the police. Attempts to solicit children have become common in chat rooms, but manual monitoring of each conversation is impossible, due to the massive amount of data and privacy issues. Therefore, development of reliable tools for detecting pedophilia in social media is of great importance.

Another related issue is that Internet makes it very easy to provide false personal information. Therefore, many online sexual predators create false profiles where they hide their identity and age. Thus, detection of online sexual predation also involves age and gender detection in chats.

From the Natural Language Processing (NLP) perspective, there are additional challenges to this problem because of the chat data specificity. Chat conversations are very different, not only from the written text, but also from other types of Internet communication, such as blogs and forums. Since online chatting usually involves very fast typing, mistakes, misspellings, and abbreviations occur frequently in chats. Moreover, specific slang (e.g. “kewl” is used instead of “cool” and “asl” stands for “age/sex/location”) and character flooding (e.g. *greeeeeat!*) are used. Therefore, modern NLP tools often fail to provide accurate processing of chat language.

Previous research on cyberpedophiles reports that they often copy juveniles’ behavior (Egan et al., 2011), in particular, they often use colloquialisms and emoticons. Other important characteristics reported previously include the unwillingness of the predator to step out of the sex-related conversation, even if the potential victim wants to change the topic. This is called fixated discourse (Egan et al., 2011). In this paper we present preliminary experiments on modelling this phenomenon. To approach the problem we apply lexical chaining techniques. The experiments show the difference in the length of sex-related lexical chains between different datasets. We believe this fact could be then utilized in detecting pedophiles.

The following section overviews related work on the topic. Section 3 briefly describes previous research on pedophiles, the language of online sexual predation and the fixated discourse phenomenon in particular. Our approach to modelling fixated discourse is presented in Section 4. We describe the data set used in the experiments in Section 5, followed by preliminary experiments presented in Section 6. We finally draw some conclusions and plans for future work in Section 7.

Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection, pages 86–90,
Avignon, France, April 23 - 27 2012.

2 Related Work

The problem of detecting pedophiles in social media is difficult and relatively novel. New ways of meeting new friends are offered: chatting with webcam (http://chatroulette.com/) or picking another user at random and let you have a one-on-one chat with each other (http://omegle.com/) in a completely anonymous way.

Some chat conversations with online sexual predators are available at www.perverted-justice.com. The site is run by adult volunteers who enter chat rooms as juveniles (usually 12-15 year old) and if they are sexually solicited by adults, they work with the police to prosecute this. Related to the problem of pedophile detection in social media, a study of Perverted Justice Foundation revealed that since 2007, they have been working on identifying sex offenders on Myspace and in 2008, they expanded that effort to Facebook. The results are sadly staggering in terms of sex offenders that have misused the two social media: Myspace (period 2007- 2010) and Facebook (2008-2010) deleted respectively 10,746 and 2,800 known sex offenders. Although both social media have been helpful and responsive towards removing danger users from their communities, an automatic identification of sex offenders would certainly help and make the process faster.

Only few attempts to automatic detection of online sexual predation have been done. Pendar (2007) proved that it is possible to distinguish between predator and pseudo-victim with quite high accuracy. The experiments were conducted on perverted-justice data. The authors used a kNN classifier to distinguish between lines written by predators and the lines posted by pseudo-victims. As features they used word unigrams, bigrams and trigrams.

Another attempt has been done by McGhee et al. (2011). They manually annotated the chat lines from perverted-justice.com with the following labels:

1. Exchange of personal information
2. Grooming
3. Approach
4. None of the above listed classes

In order to distinguish between these types of lines they used both a rule-based and a machine learning (kNN) classification approach. Their experiments showed that the machine learning approach provides better results and achieves up to 83% accuracy.

Another research work closely related to detection of cyberpedophilia has been carried by Peersman et al. (**?**). As it was already mentioned, pedophiles often create false profiles and pretend to be younger or of another gender. Moreover, they try to copy children's behaviour. Therefore, there is a need to detect age and gender in chat conversation. Peersman et al. (**?**) have analyzed chats from Belgium Netlog social network. Discrimination between those who are older than 16 from those who are younger based on Support Vector Machine classification yields 71.3% accuracy. The accuracy is even higher with increasing the gap between the age groups (e.g. the accuracy of classifying those who are less than 16 from those who are older than 25 is 88.2%). They have also investigated the issues of the minimum required dataset. Their experiments have shown that with 50% of the original dataset the accuracy remains almost the same and with only 10% it is still much better than random baseline performance.

3 Profiling the Pedophile

Pedophilia is a "disorder of adult personality and behaviour" which is characterized by sexual interest in prepubescent children (International statistical classification of diseases and related health problems, 1988). Even though solicitation of children is not a medical diagnosis, Abel and Harlow (2001) reported that 88% of child sexual abuse cases are committed by pedophiles. Therefore, we believe that understanding behaviour of pedophiles could help detecting and preventing online sexual predation. Even though online sexual offender is not always a pedophile, in this paper we use these terms as synonyms.

3.1 Predator's Linguistic Behavior

The language sexual offenders use was analyzed by Egan et al. (2011). The authors considered the chats published at www.perverted-justice.com. The analysis of the chats revealed several characteristics of predators' language:

- Fixated discourse. Predators impose a sex-related topic on the conversation and dismiss attempts from the pseudo-victim to switch topics.
- Implicit/explicit content. On the one hand, predators shift gradually to the sexual conversation, starting with more ordinary compliments. On the other hand, conversation then becomes overtly related to sex. They do not hide their intentions.
- Offenders often understand that what they are doing is not moral.
- They transfer responsibility to the victim.
- Predators often behave as children, copying the language: colloquialisms often appear in their messages.
- They try to minimize the risk of being prosecuted: they ask to delete chat logs and warn victims not to tell anyone about the talk, though they finally stop being cautious and insist on meeting offline.

In this paper we consider only the first characteristic: fixated discourse. The conversation below, taken from perverted-justice.com, illustrates fixated discourse: the predator almost ignores what the victim says and comes back to the sex-related conversation:

> **Predator**: licking dont hurt
> **Predator**: its like u lick ice cream
> **Pseudo-victim**: do u care that im 13 in march and not yet? i lied a little bit b4
> **Predator**: its all cool
> **Predator**: i can lick hard

4 Our Approach

We believe that lexical chains are appropriate to model the fixated discourse of the predators chats.

4.1 Lexical Chains

A lexical chain is a sequence of semantically related terms (Morris and Hirst, 1991). It has applications in many tasks including Word Sense Disambiguation (WSD) (Galley and McKeown, 2003) and Text Summarization (Barzilay and Elhadad, 1997).

To estimate semantic similarity we used two metrics: the similarity of Leacock and Chodorow (Leacock and Chodorow, 2003), and that of Resnik (Resnik, 1995). Leacock and Chodorow's semantic similarity measure is defined as:

$$Sim_{L\&Ch}(c_1, c_2) = -\log\frac{length(c_1, c_2)}{2 * depth}$$

where *length*(c_1, c_2) is the length of the shortest path between the concepts c_1 and c_2 and *depth* is depth of the taxonomy.

The semantic similarity measure that was proposed by Resnik (Resnik, 1995) relies on the Information Content concept:

$$IC(c) = -\log P(c)$$

where *P(c)* is the probability of encountering the concept *c* in a large corpus. Thus, Resnik's similarity measure is defined as follows:

$$Sim_{Resnik}(c_1, c_2) = IC(lcs(c_1, c_2))$$

where *lcs*(c_1, c_2) is the least common subsumer of c_1 and c_2.

4.2 Modelling Fixated Discourse

To model the fixated discourse phenomenon, we estimate the length of the longest sex-related lexical chain in a text. In particular, we start the construction of a chain with an anchor word "sex" in the first WordNet meaning: "sexual activity, sexual practice, sex, sex activity (activities associated with sexual intercourse)". Then we continue the chain construction process until the end of the text. We compare the relative lengths (in percentage to the total number of words) of the constructed chains: we believe that the presence of a long sex-related lexical chain in a text indicates fixated discourse.

5 Data

Pendar (2007) has summarized the possible types of chat interactions with sexually explicit content:

1. Predator/Other
   (a) Predator/Victim (victim is underage)
   (b) Predator/Volunteer posing as a children
   (c) Predator/Law enforcement officer posing as a child
2. Adult/Adult (consensual relationship)

The most interesting from our research point of view is data of the type 1(a), but obtaining such data is not easy. However, the data of type 1(b) is freely available at the web site www.perverted-justice.com (PJ). For our study, we have extracted chat logs from the perverted-justice website. Since the victim is not real, we considered only the chat lines written by predators.

As the negative dataset, we need data of type 2. Therefore, we have downloaded cybersex chat logs available at www.oocities.org/urgrl21f/. The archive contains 34 one-on-one cybersex logs. We have separated lines of different authors, thereby obtaining 68 files.

We have also used a subset of the NPS chat corpus (Forsythand and Martell, 2007), though it is not of type 2, we believe it will make a good comparison. We have extracted chat lines only for those adult authors who had more than 30 lines written. Finally the NPS dataset consisted of 65 authors.

6 Experiments

We carried out preliminary experiments on estimating the length of lexical chains with sexually related content in PJ chats, and compare our results with the corpora described above. Our goal is to explore the feasibility of including fixated discourse as a feature in pedophile detection.

We used Java WordNet Similarity library (Hope, 2008), which is a Java implementation of Perl Wordnet:Similarity (Pedersen et al., 2008). The average length of the longest lexical chains (with respect to the total number of words in a document) found for different corpora are presented in Table 1 and Table 2. As we expected, sex-related lexical chains in the NPS corpus are much shorter regardless of the similarity metric used. The chains in the cybersex corpus are even longer than in PJ corpus. This is probably due

	Threshold			
	0.5		0.7	
	mean	st.dev.	mean	st.dev.
PJ	12.21	3.63	9.3	5.68
Cybersex	18.28	16.8	9.98	12.76
NPS	5.66	5.9	2.42	4.77

Table 1: Average length of the longest lexical chain (percentage in the total number of words) computed with Leacock and Chodorow semantic similarity.

	Threshold			
	0.5		0.7	
	mean	st.dev.	mean	st.dev.
PJ	8.24	4.51	6.68	5.06
Cybersex	12.04	15.86	9.13	11.64
NPS	0.67	0.96	0.41	0.66

Table 2: Average length of the longest lexical chain (percentage in the total number of words) computed with Resnik semantic similarity.

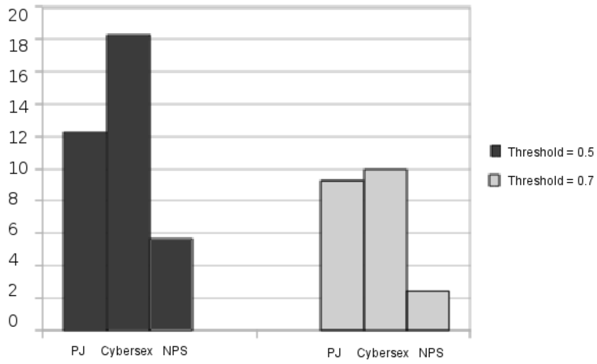

Figure 1: Average length of lexical chains calculated with Leacock and Chodorow semantic similarity

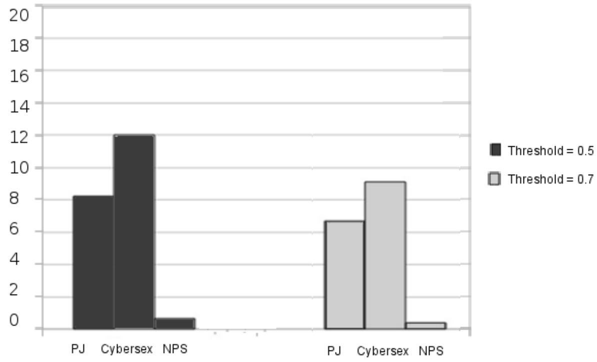

Figure 2: Average length of lexical chains calculated with Resnik semantic similarity

to the fact that whilst both corpora contain conversations about sex, cyberpedophiles are switching to this topic gradually, whereas cybersex logs are entirely sex-related.

7 Conclusions and Future Work

Detection of online sexual predation is a problem of great importance. In this small scale study we have focused on modelling fixated discourse using lexical chains as a potential feature in the automated detection of online sex predators. The preliminary experiments revealed that the lengths of sex-related lexical chains vary with the nature of the corpus, with the pedophiles logs having longer lexical chains than chat logs not related to sex, while the cybersex chat logs had the longest sex-related lexical chains of the three corpora.

As it was mentioned in Section 1, chat language is very informal and has a lot of abbreviations, slang words, mistakes etc. Hence a fair amount of words used there do not appear in WordNet and, therefore, can not be included into the lexical chains. For example, the word "ssex" is obviously related and should appear in the chain, though because of the different spelling it is not found in WordNet and, therefore, is not included into the chain. We plan to add a normalization step prior to computing lexical chains. We have used only one anchor word ("sex") to start the lexical chain. But several other words could also be good candidate for this.

Fixated discourse is not only about keeping the sexual topic throughout all the conversation, it is also about unwillingness to step out of the sexual conversation and ignoring victim's attempts to do it. Therefore, the chat lines of the pseudo-victim should be analyzed as well in order to find out if there were failed attempts to switch the topic. This may also help to distinguish predation from cybersex conversation, since in the cybersex conversation both participants want to follow the topic. However, during this preliminary experiments we have not yet considered this. Moreover, perverted-justice is run by volunteers posing as potential victims. It is then possible that the volunteers' behavior differ from the responses of real children (Egan et al., 2011). Their goal is to build a legal case against the pedophile and, therefore, they are more willing to provoke the predator than to avoid sex-related conversation.

Another way to distinguish cybersex fixed topic from the predator's unwillingness to step out of it is could be to use emotion classification based on the Leary Rose model proposed by Vaassen and Daelemans (Vaassen and Daelemans, 2011). Their approach is based on Interpersonal Circumplex suggested by Leary (Leary, 1957). This is a model of interpersonal communication that reflects whether one of the participants is dominant and whether the participants are cooperative. It was already mentioned that cyberpedophiles tend to be dominant. Therefore, we believe that the Leary Rose model can be useful in detecting online sexual predation.

Once the model of fixated discourse is improved, we plan to use it as an additional feature to detect pedophiles in social media.

Acknowledgements

The first author was partially supported by a Google Research Award and by a scholarship from the University of St. Petersburg. The second author was supported by WIQ-EI IRSES project (grant no. 269180) from the European Commission, within the FP 7 Marie Curie People, the MICINN research project TEXT-ENTERPRISE 2.0 TIN2009-13391-C04-03(Plan I+D+i), and the VLC/CAMPUS Microcluster on Multimodal Interaction in Intelligent Systems. The last author was partially supported by the UPV program PAID-02-11, award no. 1932.

References

Gene G. Abel and Nora Harlow. The Abel and Harlow child molestation prevention study. Philadelphia, Xlibris, 2001.

Regina Barzilay and Michael Elhadad. Using lexical chains for text summarization. *In Proceedings of the Intelligent Scalable Text Summarization Workshop*, 1997.

Vincent Egan, James Hoskinson, and David Shewan. Perverted justice: A content analysis of the language used by offenders detected attempting to solicit children for sex. *Antisocial Behavior: Causes, Correlations and Treatments*, 2011.

Eric N Forsythand and Craig H Martell. Lexical and discourse analysis of online chat dialog. *International Conference on Semantic Computing ICSC 2007*, pages 19–26, 2007.

Michel Galley and Kathleen McKeown. Improving word sense disambiguation in lexical chaining. *In Proceedings of IJCAI-2003*, 2003.

David Hope. Java wordnet similarity library. http://www.cogs.susx.ac.uk/users/drh21.

Claudia Leacock and Martin Chodorow. C-rater: Automated scoring of short-answer questions. *Computers and the Humanities*, 37(4):389–405, 2003.

Timothy Leary. *Interpersonal diagnosis of personality; a functional theory and methodology for personality evaluation*. Oxford, England: Ronald Press, 1957.

India McGhee, Jennifer Bayzick, April Kontostathis, Lynne Edwards, Alexandra McBride and Emma Jakubowski. Learning to identify Internet sexual predation. *International Journal on Electronic Commerce* 2011.

Kimberly J. Mitchell, David Finkelhor, and Janis Wolak. Risk factors for and impact of online sexual solicitation of youth. *Journal of the American Medical Association*, 285:3011–3014, 2001.

Jane Morris and Graeme Hirst. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1):21–43, 1991.

Federal Bureau of Investigation. Nibrs flatfile tape master record descriptions. 1995.

Ted Pedersen, Siddharth Patwardhan, Jason Michelizzi, and Satanjeev Banerjee. Wordnet:similarity. http://wn-similarity.sourceforge.net/.

Nick Pendar. Toward spotting the pedophile: Telling victim from predator in text chats. pages 235–241, Irvine, California, 2007.

Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *IJCAI*, pages 448–453, 1995.

Howard N. Snyder. Sexual assault of young children as reported to law enforcement: Victim, incident, and offender characteristics. a nibrs statistical report. *Bureau of Justice Statistics Clearinghouse*, 2000.

Frederik Vaassen and Walter Daelemans. Automatic emotion classification for interpersonal communication. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 104–110. Association for Computational Linguistics, 2011.

World health organization, international statistical classification of diseases and related health problems: Icd-10 section f65.4: Paedophilia. 1988.