

On the Impact of Sentiment and Emotion Based Features in Detecting Online Sexual Predators

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

Paolo Rosso
NLE Lab - ELiRF
Universitat
Politècnica de València
prossor@dsic.upv.es

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

Abstract

According to previous work on pedophile psychology and cyberpedophilia, sentiments and emotions in texts could be a good clue to detect online sexual predation. In this paper, we have suggested a list of high-level features, including sentiment and emotion based ones, for detection of online sexual predation. In particular, since pedophiles are known to be emotionally unstable, we were interested in investigating if emotion-based features could help in their detection. We have used a corpus of predators' chats with pseudo-victims downloaded from www.perverted-justice.com and two negative datasets of different nature: cybersex logs available online and the NPS chat corpus. Naive Bayes classification based on the proposed features achieves accuracies of up to 94% while baseline systems of word and character n-grams can only reach up to 72%.

1 Introduction

Child sexual abuse and pedophilia are both problems 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 the pedophilia has been studied from different research points, it remains to be a very important problem which requires further research, especially from the automatic detection point of view.

Previous studies report that in the majority of cases of sexual assaults the victims are underage (Snyder, 2000). On the Internet, attempts

to solicit children have become common as well. Mitchell (2001) found out that 19% of children have been sexually approached online. However, manual monitoring of each conversation is impossible, due to the massive amount of data and privacy issues. A good alternative is the development of reliable tools for detecting pedophilia in online social media is of great importance.

In this paper, we address the problem of detecting pedophiles with natural language processing (NLP) techniques. This problem becomes even more challenging because of the chat data specificity. Chat conversations are very different not only from the written text but also from other types of social media interactions, such as blogs and forums, since chatting in the Internet usually involves very fast typing. The data usually contains a large amount of mistakes, misspellings, specific slang, character flooding etc. Therefore, accurate processing of this data with automated syntactic analyzers is rather challenging.

Previous research on pedophilia reports that the expression of certain emotions in text could be helpful to detect pedophiles in social media (Egan et al., 2011). Following these insights we suggest a list of features, including sentiments as well as other content-based features. We investigate the impact of these features on the problem of automatic detection of online sexual predation. Our experimental results show that classification based on such features discriminates pedophiles from non-pedophiles with high accuracy.

The remainder of the paper is structured as follows: Section 2 overviews related work on the topic,

Section 3 outlines the profile of a pedophile based on the previous research. Our approach to the problem of detecting pedophiles in social media on the basis of high-level features is presented in Section 4. Experimental data is described in Section 5. We show the results of the conducted experiments in Section 6; they are followed by discussion and plans for future research in Section 7. We finally draw some conclusions in Section 8.

2 Related Research

The problem of automatic detection of pedophiles in social media has been rarely addressed so far. In part, this is due to the difficulties involved in having access to useful data. There is an American foundation called Perverted Justice (PJ). It investigates cases of online sexual predation: adult volunteers 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 the offenders. Some chat conversations with online sexual predators are available at www.perverted-justice.com and they have been the subject of analysis of recent research on this topic.

Pendar (2007) experimented with PJ data. He separated the lines written by pedophiles from those written by pseudo-victims and used a kNN classifier based on word n-grams to distinguish between them.

Another related research has been carried out by McGhee et al. (2011). The chat lines from PJ were manually classified into the following categories:

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

Their experiments have shown that kNN classification achieves up to 83% accuracy and outperforms a rule-based approach.

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 behavior. Automatically detecting age and gender in chat conversations could then be the

first step in detecting online predators. Peersman et al. (2011) have analyzed chats from the Belgium Netlog social network. Discrimination between those who are older than 16 from those who are younger based on a Support Vector Machine classification yields 71.3% accuracy. The accuracy is even higher when the age gap is increased (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 amount of training data needed. 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 the random baseline performance.

NLP techniques were as well applied to capture child sexual abuse data in P2P networks (Panchenko et al., 2012). The proposed text classification system is able to predict with high accuracy if a file contains child pornography by analyzing its name and textual description.

Our work neither aims at classification of chat lines into categories as it was done by McGhee et al. (2011) nor at discriminating between victim and predator as it was done by Pendar (2007), but at distinguishing between pedophile's and not pedophile's chats, in particular, by utilizing clues provided by psychology and sentiment analysis.

3 Profiling the Pedophile

Pedophilia is a "disorder of adult personality and behavior" 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 behavior of pedophiles could help to detect and prevent online sexual predation. Even though an online sexual offender is not always a pedophile, in this paper we use these terms as synonyms.

Previous research reports that about 94% of sexual offenders are males. With respect to female sexual molesters, it is reported, that they tend to be young and, in these cases, men are often involved as well (Vandiver and Kercher, 2004). Sexual as-

sault offenders are more often adults (77%), though in 23% of cases children are solicited by other juveniles.

Analysis of pedophiles' personality characterizes them with feelings of inferiority, isolation, loneliness, low self-esteem and emotional immaturity. Moreover, 60%-80% of them suffer from other psychiatric illnesses (Hall and Hall, 2007). In general, pedophiles are less emotionally stable than mentally healthy people.

3.1 Profile of the Online Sexual Predator

Hall and Hall (2007) noticed that five main types of computer-based sexual offenders can be distinguished: (1) the stalkers, who approach children in chat rooms in order to get physical access to them; (2) the cruisers, who are interested in online sexual molestation and not willing to meet children offline; (3) the masturbators, who watch child pornography; (4) the networkers or swappers, who trade information, pornography, and children; and (5) a combination of the four types. In this study we are interested in detecting stalkers (type (1)) and cruisers (type (2)).

The language sexual offenders use was analyzed by Egan et al. (2011). The authors considered the chats available from PJ. The analysis of the chats revealed several characteristics of predators' language:

- Implicit/explicit content. On the one hand, predators shift gradually to the sexual conversation, starting with more ordinary compliments:

Predator: hey you are really cute

Predator: u are pretty

Predator: hi sexy

On the other hand, the conversation then becomes overtly related to sex. They do not hide their intentions:

Predator: can we have sex?

Predator: you ok with sex with me and drinking?

- Fixated discourse. Predators are not willing to step aside from the sexual conversation. For example, in this conversation the predator almost ignores the question of pseudo-victim 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

- Offenders often understand that what they are doing is not moral:

Predator: i would help but its not moral

- They transfer responsibility to the victim:

Pseudo-victim: what ya wanta do when u come over

Predator: whatever—movies, games, drink, play around—it's up to you—what would you like to do?

Pseudo-victim: that all sounds good

Pseudo-victim: lol

Predator: maybe get some sexy pics of you :-P

Predator: would you let me take pictures of you? of you naked? of me and you playing? :-D

- Predators often behave as children, copying their linguistic style. Colloquialisms appear often in their messages:

Predator: howwwwww dy

...

Predator: i know PITY MEEEE

- 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:

Predator: don't tell anyone we have been talking

Pseudo-victim: k

Pseudo-victim: lol who would i tell? no one's here.

Predator: well I want it to be our secret

- Though they finally stop being cautious and insist on meeting offline:

Predator: well let me come see you

Pseudo-victim: why u want 2 come over so bad?

Predator: i wanna see you

In general Egan et al. (Egan et al., 2011) have found online solicitation to be more direct, while in real life children seduction is more deceitful.

4 Our Approach

We address the problem of automatic detection of online sexual predation. While previous studies were focused on classifying chat lines into different categories (McGhee et al., 2011) or distinguishing between offender and victim (Pendar, 2007), in this work we address the problem of detecting sexual predators.

We formulate the problem of detecting pedophiles in social media as the task of binary text categorization: given a text (a set of chat lines), the aim is to predict whether it is a case of cyberpedophilia or not.

4.1 Features

On the basis of previous analysis of pedophiles' personality (described in previous section), we consider as features those emotional markers that could unveil a certain degree of emotional instability, such as feelings of inferiority, isolation, loneliness, low self-esteem and emotional immaturity.

On the one hand, pedophiles try to be nice with a victim and make compliments, at least in the beginning of a conversation. Therefore, the use of positive words is expected. On the other hand, as it was described earlier, pedophiles tend to be emotionally unstable and prone to lose temper, hence they might

start using words expressing anger and negative lexicon. Other emotions can be as well a clue to detect pedophiles. For example, offenders often demonstrate fear, especially with respect to being prosecuted, and they often lose temper and express anger:

Pseudo-victim: u sad didnt car if im 13. now u car.

Predator: well, *I am just scared* about being in trouble or going to jail

Pseudo-victim: u sad run away now u say no. i gues i dont no what u doin

Predator: *I got scared*

Predator: we would get caught sometime

In this example pseudo-victim is not answering:

Predator: hello

Predator: r u there

Predator:

Predator: thnx a lot

Predator: thanx a lot

Predator:

Predator: *u just wast my time*

Predator: drive down there

Predator: can u not im any more

Here the offender is angry because the pseudo-victim did not call him:

Predator: u didnt call

Predator: *i m angry with u*

Therefore, we have decided to use markers of basic emotions as features. At the SemEval 2007 task on "Affective Text" (Strapparava and Mihalcea, 2007) the problem of fine-grained emotion annotation was defined: given a set of news titles, the system is to label each title with the appropriate emotion out of the following list: **ANGER, DISGUST, FEAR, JOY, SADNESS, SURPRISE**. In this research work we only use the percentages of the markers of each emotion.

We have also borrowed several features from McGhee et al. (2011):

- **Percentage of approach words.** Approach words include verbs such as *come* and *meet* and such nouns as *car* and *hotel*.
- **Percentage of relationship words.** These words refer to dating (e.g. *boyfriend*, *date*).

- **Percentage of family words.** These words are the names of family members (e.g. *mum, dad, brother*).
- **Percentage of communicative desensitization words.** These are explicit sexual terms offenders use in order to desensitize the victim (e.g. *penis, sex*).
- **Percentage of words expressing sharing information.** This implies sharing basic information, such as age, gender and location, and sending photos. The words include *asl, pic*.

Since pedophiles are known to be emotionally unstable and suffer from psychological problems, we consider features reported to be helpful to detect neuroticism level by Argamon et al. (2009). In particular, the features include *percentages of personal* and *reflexive pronouns* and *modal obligation verbs* (have to, has to, had to, must, should, mustn't, and shouldn't).

We consider the use of imperative sentences and emoticons to capture the predators tendencies to be dominant and copy childrens' behaviour respectively.

The study of Egan et al. (Egan et al., 2011) has revealed several recurrent themes that appear in PJ chats. Among them, *fixated discourse*: the unwillingness of the predator to change the topic. In (Bogdanova et al., 2012) we present experiments on modeling the fixated discourse. We have constructed lexical chains (Morris and Hirst, 1991) starting with the anchor word "sex" in the first WordNet meaning: "sexual activity, sexual practice, sex, sex activity (activities associated with sexual intercourse)". We have finally used as a feature the *length of the lexical chain* constructed with the Resnik similarity measure (Resnik, 1995) with the threshold = 0.7.

The full list of features is presented in Table 1.

5 Datasets

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

1. Predator/Other
 - (a) Predator/Victim (victim is underaged)
 - (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 1a, but obtaining such data is not easy. However, the data of the type 1b is freely available at the web site www.perverted-justice.com. 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.

Since our goal is to distinguish sex related chat conversations where one of the parties involved is a pedophile, the ideal negative dataset would be chat conversations of type 2 (consensual relations among adults) and the PJ data will not meet this condition for the negative instances. We need additional chat logs to build the negative dataset. We used two negative datasets in our experiments: cybersex chat logs and the NPS chat corpus.

We downloaded the cybersex chat logs available at www.oocities.org/urgr121f/. The archive contains 34 one-on-one cybersex logs. We have separated lines of different authors, thereby obtaining 68 files.

We have also used the subset the of NPS chat corpus (Forsythand and Martell, 2007), though it is not of type 2. We have extracted chat lines only for those adult authors who had more than 30 lines written. Finally the dataset consisted of 65 authors. From each dataset we have left 20 files for testing.

6 Experiments

To distinguish between predators and not predators we used a Naive Bayes classifier, already successfully utilized for analyzing chats by previous research (Lin, 2007). To extract positive and negative words, we used SentiWordNet (Baccianella et al., 2010). The features borrowed from McGhee et al. (2011), were detected with the list of words authors made available for us. Imperative sentences were detected as affirmative sentences starting with verbs. Emoticons were captured with simple regular expressions.

Our dataset is imbalanced, the majority of the chat logs are from PJ. To make the experimental data more balanced, we have created 5 subsets of PJ cor-

Feature Class	Feature	Example	Resource
Emotional Markers	Positive Words	<i>cute, pretty</i>	SentiWordNet (Baccianella et al., 2010)
	Negative Words	<i>dangerous, annoying</i>	
	JOY words	<i>happy, cheer</i>	WordNet-Affect (Strapparava and Valitutti, 2004)
	SADNESS words	<i>bored, sad</i>	
	ANGER words	<i>annoying, furious</i>	
	SURPRISE words	<i>astonished, wonder</i>	
	DISGUST words	<i>yucky, nausea</i>	
	FEAR words	<i>scared, panic</i>	
Features borrowed from McGhee et al. (2011)	Approach words	<i>meet, car</i>	McGhee et al. (2011)
	Relationship nouns	<i>boyfriend, date</i>	
	Family words	<i>mum, dad</i>	
	Communicative desensitization words	<i>sex, penis</i>	
Features helpful to detect neuroticism level	Information words	<i>asl, home</i>	Argamon et al. (2009)
	Personal pronouns	<i>I, you</i>	
	Reflexive pronouns	<i>myself, yourself</i>	
	Obligation verbs	<i>must, have to</i>	
Features derived from pedophile's psychological profile	Fixated Discourse	see in Section 3.1	Bogdanova et al. (2012)
Other	Emoticons	8), :(
	Imperative sentences	<i>Do it!</i>	

Table 1: Features used in the experiments.

pus, each of which contained chat lines from 60 randomly selected predators.

For the cybersex logs, half of the chat sessions belong to the same author. We used this author for training, and the rest for testing, in order to prevent the classification algorithm from learning to distinguish this author from pedophiles.

For comparison purposes, we experimented with several baseline systems using low-level features based on n-grams at the word and character level, which were reported as useful features by related research (Peersman et al., 2011). We trained naive Bayes classifiers using word level unigrams, bigrams and trigrams. We also trained naive Bayes classifiers using character level bigrams and trigrams.

The classification results are presented in Tables 2 and 3. The high-level features outperform all the low-level ones in both the cybersex logs and the NPS chat datasets and achieve 94% and 90% accuracy on these datasets respectively.

Cybersex chat logs are data of type 2 (see previous section), they contain sexual content and, therefore, share same of the same vocabulary with the

perverted-justice data, whilst the NPS data generally is not sex-related. Therefore, we expected low-level features to provide better results on the NPS data. The experiments have shown that, except for the character bigrams, all low-level features considered indeed work worse in case of cybersex logs (see the average rows in both tables). The average accuracy in this case varies between 48% and 58%. Surprisingly, low-level features do not work as good as we expected in case of the NPS chat dataset: bag of words provides only 61% accuracy. Among other low-level features, character trigrams provide the highest accuracy of 72%, which is still much lower than the one of the high-level features (90%). The high-level features yield a lower accuracy (90% accuracy) on the PJ-NPS dataset than in the case of PJ-cybersex logs (94% accuracy). This is probably due to the data diversity: cybersex chat is a very particular type of a conversation, though NPS chat corpora can contain any type of conversations up to sexual predation.

	Accuracy					
	High-level features	Bag of words	Term bigrams	Term trigrams	Character bigrams	Character trigrams
Run 1	0.93	0.38	0.55	0.60	0.73	0.78
Run 2	0.95	0.40	0.50	0.53	0.75	0.45
Run 3	0.95	0.70	0.45	0.53	0.48	0.50
Run 4	0.98	0.43	0.53	0.53	0.50	0.38
Run 5	0.90	0.50	0.48	0.53	0.45	0.50
Average	0.94	0.48	0.50	0.54	0.58	0.52

Table 2: Results of Naive Bayes classification applied to perverted-justice data and cybersex chat logs.

	Accuracy					
	High-level features	Bag of words	Term bigrams	Term trigrams	Character bigrams	Character trigrams
Run 1	0.93	0.73	0.60	0.60	0.68	0.75
Run 2	0.95	0.68	0.53	0.53	0.48	0.45
Run 3	0.95	0.58	0.53	0.53	0.48	0.85
Run 4	0.98	0.53	0.53	0.53	0.23	0.80
Run 5	0.90	0.53	0.53	0.53	0.25	0.75
Average	0.92	0.61	0.54	0.54	0.42	0.72

Table 3: Results of Naive Bayes classification applied to perverted-justice data and NPS chats.

7 Discussion and Future Work

We have conducted experiments on detecting pedophiles in social media with a binary classification algorithm. In the experiments we used two negative datasets of different nature: the first one is more appropriate, it contains one-on-one cybersex conversations, while the second dataset is extracted from the NPS chat corpus and contains logs from chat rooms, and, therefore, is less appropriate since the conversations are not even one on one.

It is reasonable to expect that in the case of the negative data consisting of cybersex logs, distinguishing cyberpedophiles is a harder task, than in the case of the NPS data. The results obtained with the baseline systems support this assumption: we obtain higher accuracy for the NPS chats in all but character bi-grams. The interesting insight from these results is that our proposed higher-level features are able to boost accuracy to 94% on the seemingly more challenging task.

Our error analysis showed that the NPS logs misclassified with the high-level features are also misclassified by the baseline systems. These instances either share the same lexicon or are about the same topics. Therefore they are more similar to cyberpe-

dophiles training data than the training data of the NPS corpus, which is very diverse. These examples are taken from misclassified NPS chat logs:

User: love me like a bomb baby come on get it on

...

User: ryaon so sexy

User: you are so anal

User: obviously i didn't get it

User: just loosen up babe

...

User: i want to make love to him

User: right field wrong park lol j/k

User: not me i put them in the jail lol

User: or at least tell the cops where to go to get the bad guys lol

In the future we plan to further investigate the misclassified data. The feature extraction we have implemented does not use any word sense disambiguation. This can as well cause mistakes since the markers are not just lemmas but words in particular senses, since for example the lemma “fit” can be either a positive marker (“a fit candidate”) or negative (“a fit of epilepsy”), depending on the

context. Therefore we plan to employ word sense disambiguation techniques during the feature extraction phase.

So far we have only seen that the list of features we have suggested provides good results. They outperform all the low-level features considered. Among those low-level features, character trigrams provide the best results on the NPS data (72% accuracy), though on the cybersex logs they achieve only 54%. We plan to merge low-level and high-level features in order to see if this could improve the results.

In the future we plan also to explore the impact of each high-level feature. To better understand which ones carry more discriminative power and if we can reduce the number of features. All these experiments will be done employing naive Bayes as well as Support Vector Machines as classifiers.

8 Conclusions

This paper presents some results of an ongoing research project on the detection of online sexual predation, a problem the research community is interested in, as the PAN task on Sexual Predator Identification¹ suggests.

Following the clues given by psychological research, we have suggested a list of high-level features that should take into account the level of emotional instability of pedophiles, as well as their feelings of inferiority, isolation, loneliness, low self-esteem etc. We have considered as well such low-level features as character bigrams and trigrams and word unigrams, bigrams and trigrams. The Naive Bayes classification based on high-level features achieves 90% and 94% accuracy when using NPS chat corpus and the cybersex chat logs as a negative dataset respectively, whereas low-level features achieve only 42%-72% and 48%-58% accuracy on the same data.

Acknowledgements

The research of Dasha Bogdanova was carried out during the 3-month internship at the Universitat Politècnica de València (scholarship of the University of St.Petersburg). Her research was partially

supported by Google Research Award. The collaboration with Tamar Solorio was possible thanks to her one-month research visit at the Universitat Politècnica de València (program PAID-PAID-02-11 award n. 1932). The research work of Paolo Rosso was done in the framework of the European Commission WIQ-EI IRSES project (grant no. 269180) 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.

References

- Gene G. Abel and Nora Harlow. The Abel and Harlow child molestation prevention study. Philadelphia, Xlibris, 2001.
- Shlomo Argamon, Moshe Koppel, James Pennebaker, and Jonathan Schler. Automatically profiling the author of an anonymous text. *Communications of the ACM*, 52 (2):119–123, 2009.
- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *the Seventh International conference on Language Resources and Evaluation*, 2010.
- Regina Barzilay and Michael Elhadad. Using lexical chains for text summarization. *In Proceedings of the Intelligent Scalable Text Summarization Workshop*, 1997.
- Dasha Bogdanova, Paolo Rosso, Tamar Solorio. Modelling Fixated Discourse in Chats with Cyberpedophiles. *Proceedings of the Workshop on Computational Approaches to Deception Detection, EACL*, 2012.
- 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 Forsyth 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.
- Ryan C. W. Hall and Richard C. W. Hall. A profile of pedophilia: Definition, characteristics of offenders, recidivism, treatment outcomes, and forensic issues. *Mayo Clinic Proceedings*, 2007.

¹<http://pan.webis.de/>

- 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.
- Jane Lin. *Automatic author profiling of online chat logs*. PhD thesis, 2007.
- 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.
- Ted Pedersen, Siddharth Patwardhan, Jason Michelizzi, and Satanjeev Banerjee. Wordnet:similarity. <http://wn-similarity.sourceforge.net/>.
- Claudia Peersman, Walter Daelemans, and Leona Van Vaerenbergh. Predicting age and gender in online social networks. In *Proceedings of the 3rd Workshop on Search and Mining User-Generated Contents*, 2011.
- Nick Pendar. Toward spotting the pedophile: Telling victim from predator in text chats. In *Proceedings of the International Conference on Semantic Computing*, pages 235–241, Irvine, California, 2007.
- Alexander Panchenko, Richard Beaufort, Cedrick Fairon. Detection of Child Sexual Abuse Media on P2P Networks: Normalization and Classification of Associated Filenames. In *Proceedings of the LREC Workshop on Language Resources for Public Security Applications*, 2012.
- 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.
- Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval'07*, pages 70–74, 2007.
- Carlo Strapparava and Alessandro Valitutti. Wordnet-affect: an affective extension of wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 2004.
- 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.
- Donna M. Vandiver and Glen Kercher. Offender and victim characteristics of registered female sexual offenders in Texas: A proposed typology of female sexual offenders. *Sex Abuse*, 16:121–137, 2004.
- World health organization, international statistical classification of diseases and related health problems: Icd-10 section f65.4: Paedophilia. 1988.