Data Mining Trends and Applications in Criminal Science and Investigations

Omowunmi E. Isafiade
University of Cape Town, South Africa

Antoine B. Bagula
University of the Western Cape, South Africa

A volume in the Advances in Data Mining and Database Management (ADMDM) Book Series

Information Science REFERENCE
An Imprint of IGI Global

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Names: Isafiade, Omowunmi E., 1982- editor. | Bagula, Antoine B., 1960- editor.
Title: Data mining trends and applications in criminal science and investigations / Omowunmi E. Isafiade and Antoine B. Bagula, editors.
Description: Hershey : Information Science Reference, 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2016010961| ISBN 9781522504634 (hardcover) | ISBN 9781522504641 (ebook)
Subjects: LCSH: Criminal investigation--Data processing. | Data mining.
Classification: LCC HV7431 .D364 2016 | DDC 363.250285/6312--dc23 LC record available at https://lccn.loc.gov/2016010961

This book is published in the IGI Global book series Advances in Data Mining and Database Management (ADMDM) (ISSN: 2327-1981; eISSN: 2327-199X)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 9
Automated Identification of Child Abuse in Chat Rooms by Using Data Mining

Mohammadreza Keyvanpour
Alzahra University, Iran

Necmiye Genc Nayebi
École de Technologie Supérieure (ÉTS), Canada

Mohammadreza Ebrahimi
Concordia University, Canada

Olga Ormandjieva
Concordia University, Canada

Ching Y. Suen
Concordia University, Canada

ABSTRACT

Providing a safe environment for juveniles and children in online social networks is considered as one of the major factors of improving public safety. Due to the prevalence of the online conversations, mitigating the undesirable effects of child abuse in cyber space has become inevitable. Using automatic ways to combat this kind of crime is challenging and demands efficient and scalable data mining techniques. The problem can be casted as a combination of textual preprocessing in data/text mining and pattern classification in machine learning. This chapter covers different data mining methods including preprocessing, feature extraction and the popular ways of feature enrichment through extracting sentiments and emotional features. A brief tutorial on classification algorithms in the domain of automated predator identification is also presented through the chapter. Finally, the discussion is summarized and the challenges and open issues in this application domain are discussed.

DOI: 10.4018/978-1-5225-0463-4.ch009

1. INTRODUCTION

The ease of access and anonymity of the Internet users facilitate child exploitation, online bullying and cyber sexual abuse. This has been a major concern in developed countries with a high rate of Internet access in which children are basically the most vulnerable Internet stakeholders. Automated Online Predator Identification (OPI) is a proactive way to counteract the undesirable effects causing by aforementioned crimes. Recently in the literature, this is also known as Sexual Predator Identification (SPI) or Sexual Predator Detection (SPD). Although practical OPI problem involves dealing with textual data and images, textual data are considerably more convenient to be used for automation purposes rather than the imagery data. Accordingly, dealing with textual data is meant throughout the chapter, wherever the OPI is mentioned in general. This part highlights the importance of Online Predator Identification as an effective action toward improving public safety in society. The target audience of this chapter is the researchers working in the area of crime data mining and readers who want to have an overall grasp of the OPI field. In sections 1-1 and 1-2, we present the key concepts about public safety, OPI domain and the domain related aspects, whereas Section 1-3 will provide further information in regards to the relationship between data mining and OPI.

1.1 Public Safety and OPI

The ease of access and anonymity of the Internet users has made child abuse easier than the past. According to Kierkegaard (2008), sexual solicitations of 89% of youth are made in chat rooms. This implies the vital need for mining large volumes of anonymous chat logs in order to combat this kind of social crime. Providing a safe environment for juveniles and children in online social networks is considered as one of the major factors of improving public safety. Due to the prevalence of the online conversations, mitigating the undesirable effects of juvenile abuse in cyber space has become critical.

1.2 Domain Concepts

This part contains the essential information about legal and psychological aspects of online predator identification.

1.2.1 Legal Aspects

Although all of the legislative and regulatory provisions regarding online child sexual abuse aim to combat and mitigate the impact of this threat, they may vary in different countries or even for different jurisdictions in the same country. According

to Kierkegaard (2008), while virtual child porn using avatars is generally considered illegal in the European Union, it might not necessarily be treated as such in the United States *(p. 44)*. Similarly, *images which are illegal to view in the USA may not be illegal to view in Germany* (p. 41). The same situation exists in the concept of age disparity between the adult and minor. Hence, there have been countless writings on legal aspects of child sexual abuse in online environment which go beyond the scope of this chapter.

1.2.2 Psychological Aspects

The most effective and also naïve psychological aspects of "predatorhood" might be those defined by Morris in his master of science thesis (Morris, 2013). The author defined predatorhood as having two major components: *age disparity* and *inappropriate intimacy*. The former relates to the psychological immaturity of the victim compared to that of predator (adult) which may differ in various countries by law. The latter corresponds to the attempt of adult to establish an intimate conversation that usually involves implicit or explicit sexual comments.

One of the most practical psychological theories which is widely used in online predator identification is known as *luring communication theory* (Olson, Daggs, Ellevold, & Rogers, 2007). The theory comprises three main phases needed for committing a predatory act:

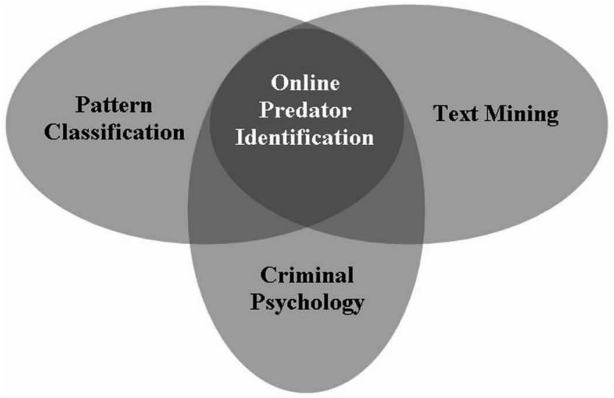
1. Gaining access to the victim.
2. Entrapping and grooming until the victim accepts sexual advances.
3. Initiating and maintaining the abusive relationship.

On the Internet, the most common way for gaining access to the victim is through online conversations in chat rooms. The second stage can be distinguished by observation of the predator's attempt to desensitize the child to inappropriate intimacy. Finally, the third step involves explicit sexual exploitation of the minor. When an explicit exploitation is about to occur, a reliable OPI system can flag the conversation for the attention of law enforcement in order to prevent the predator from approaching the victim.

1.3 Data Mining and OPI

During the past decade, automated Online Predator Identification (OPI) has become tractable by using data/text mining algorithms. These algorithms are capable of identifying likely predators for the attention of law enforcement. Using automatic ways to combat this kind of crime is challenging and demands efficient and scalable

Figure 1. The relationship of OPI with text mining, pattern classification and criminal psychology



data mining techniques that are able to handle large volumes of chat logs. Since huge amounts of text data (e.g. chat logs, web sites, and forums) are required to be processed, the automated predator identification task relies on scalable data mining algorithms. There are two major OPI problems in which data mining plays an important role:

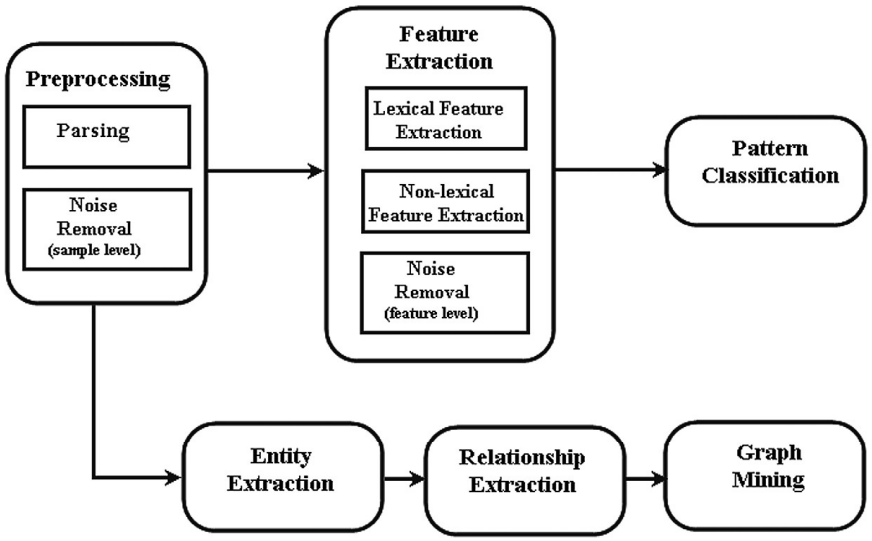
1. Detecting predators.
2. Visualizing and analyzing predator criminal networks.

The solution to these problems can be casted as a combination of textual preprocessing in data/text mining and pattern classification in machine learning and also criminal psychology (Figure 1).

As it has been illustrated in this figure, text mining and pattern classification techniques form the algorithmic foundations of OPI. A comprehensive but concise survey of text mining which incorporates the gist of text mining algorithms has been provided by Aggarwal (2015). Pattern classification (also known as pattern recognition) encompasses the classification algorithms used in predator identification and it goes hand in hand with machine learning techniques (Duda, Hart, & Stork, 2012).

More specifically, the solution to the second OPI problem is provided by extracting the underlying relationships and using graph mining techniques to analyze the resultant social networks. This topic will be discussed briefly in section 2. Figure 2 illustrates the main data mining techniques used in OPI along with their relationships. The details of each part will be given in sections 3, 4 and 5.

Figure 2. Major data mining techniques which are used in OPI and their relationships with each other



There has been a dedicated competition for Sexual Predator Identification in PAN-2012 as part of the CLEF 2012 competition ("PAN-2012," n.d.) that expedited the movement of applying data mining techniques on chat logs in order to identify the likely predators. Several competitors from all over the world applied their data mining techniques on a relatively huge volume of chat logs. The chapter covers the gist of this competition with a focus on data mining techniques. The competition encompasses the following two tasks (Inches & Crestani, 2012):

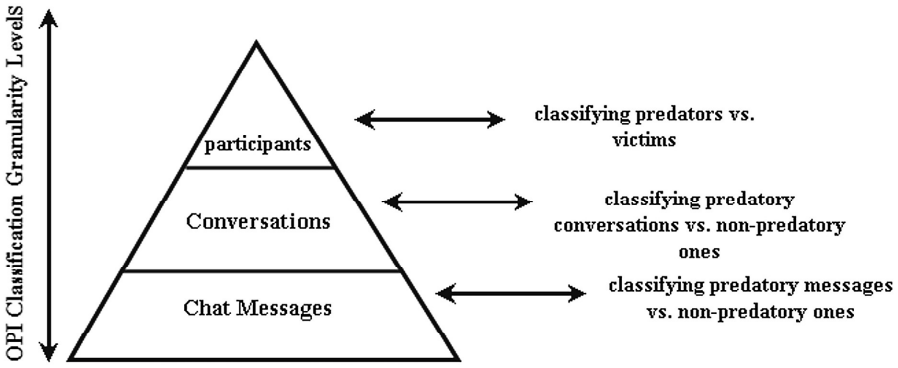
1. Distinguishing the predators and victims.
2. Specifying predatory messages in predatory conversation.

Accomplishing the first task is of greater help to law enforcement in terms of narrowing down their search space significantly. According to Villatoro-Tello et al. (2012), this task can be performed in two consecutive steps:

1. Identifying the predatory conversations among all conversations.
2. Distinguishing the sexual predator and victim among participants of predatory conversations.

We propose an abstract taxonomy which encompasses the different classification techniques which are used in the OPI field. There are three main different granularity levels of analysis in dealing with online predator identification. These levels are shown in Figure 3.

Figure 3. Classification granularity levels and their corresponding classification problem in OPI



As seen in Figure 3, the most fine-grained level of analysis corresponds to direct analysis of *messages* (also known as *interventions*) exchanged by participants in their conversation. This kind of analysis corresponds to the task of distinguishing the sexual predators. The next level of granularity relates to the task of identifying online predatory conversations. Finally, distinguishing the predators among all of the participants in the entire corpus can be considered as the highest level of abstraction which might be the ultimate goal. These analyses are accomplished by utilizing the proper data mining classification methods which are discussed in the very last sections of this chapter.

This chapter aims to cover the state-of-the-art methods and results of domain-specific papers in the field. In this regard, the usage of data mining techniques including preprocessing, feature extraction and classification have been discussed in various aspects of an OPI problem. The chapter starts with a brief introduction of terminology and a discussion on the necessity of OPI in online safety. Next, different methods of preprocessing, which have been found to be quite useful for filtering the noise and unnecessary data and improving the overall performance, are discussed. These methods include the most popular data formats, noise removal procedure and dimensionality reduction. Then feature extraction and the popular ways of feature enrichment in this domain of application are described. After discussing related feature extraction techniques, a brief basic tutorial about classification algorithms in the domain of automated predator identification is presented. These algorithms cover a wide range of classification algorithms such as Entropy-based classification, Naïve Bayes, Support Vector Machine, and Neural Networks (see section 5 for a discussion on these algorithms). Dissecting the unnecessary technical details about the internal mechanism of machine learning algorithms has been avoided. Instead, we have tried to refer the keen readers to the related resources about the fully-detailed theories behind these algorithms. Finally, we compare the performance

of these methods and introduce the most successful ones based on the outcome of CLEF 2012 conference. Finally, in the final part of the chapter we describe the open issues and remaining challenges in this field. Particularly, we address the usage of Web-based Dynamic Social Networks (WDSN) and its relationship to the online predator identification. In addition, the most likely future trends in this field and its connection with natural language processing are discussed at the end of this chapter.

2. BACKGROUND

As mentioned in the previous section, OPI has its root in text mining and pattern classification. With the rapid increase of available textual data in different domains including news, social media, and web pages, text mining has drawn the attention of researchers during the last decade. There has been a variety of algorithms and approaches including text clustering and classification, text summarization, topic modeling, and opinion mining. Opinion mining or sentiment analysis is an important discipline in Natural Language Processing (NLP) that extracts people's opinion, attitudes and emotions toward entities, other people, events, and their attributes. The reader may refer to (Aggarwal, 2015; Irfan et al., 2015; Liu & Zhang, 2012) for comprehensive explanations of these algorithms and approaches. Here, we narrow down the focus of the chapter to the usage of these techniques in OPI and describe the related background.

One of the very first successful attempts for using data mining in OPI problem was accomplished by Pendar (2007) who used weighted K-NN classifier to distinguish predators from underage victims. In addition to this, the first empirical system with capability of determining predatory messages in chat logs is ChatCoder1 (Kontostathis, 2009). Afterwards, ChatCoder2 (see section 2) was implemented and evolved on top of the previous version to improve its performance (Mcghee et al., 2011). The system used a rule based approach in conjunction with decision trees and instance-based learning methods (K-NN).

Michalopoulos and Mavridis (2011), and Escalante et al. (2013) have utilized Luring Communication Theory which was described in the previous section, to combine psychological aspects of predation phenomenon with computer science and machine learning.

Recently, the PAN-2012 conference has acted as a boost for applying machine learning techniques to this area. Several machine learning algorithms have been used to solve OPI problem in this competition. These algorithms cover a wide range of classification algorithms such as Entropy-based Classification (Eriksson & Karlgren, 2012), K-Nearest Neighbor (Kang, Kim, Kang, & Na, 2012), Support Vector Machine (Morris, 2013) and Neural Networks (Villatoro-Tello et al., 2012).

Eventually, one team has been announced as the winner based on their classification accuracy and an augmented F-measure. The winner team (Villatoro-Tello et al., 2012) used a two-step binary classification approach called SCI (Suspicious Conversation Identification) and VFP (Victim From Predator Disclosure) using SVM and Neural Networks.

As one of the most recent works, Escalante et al. (2013) proposed a novel method using the chained-classifiers based on adapting a psychological hypothesis that underscores three stages employed by predators to approach the victim. Although this method could not outperform the approach used by Villatoro-Tello et.al, it revealed that adopting psycho-linguistic hypotheses could improve the accuracy.

Due to the inadequacy of bag-of-words models in reflecting deep semantic notions hidden in the conversations, Bogdanova et al. (2012b) tried to enrich the features by introducing sentiments and emotions to the original feature set. In another research (Bogdanova, Rosso, & Solorio, 2014) they improved their feature set by adding more high level features such as neuroticism and psychological aspects. We will cover these studies in more detail through the chapter.

Deeper linguistic attempts in this field were started by Forsyth & Martell (2007) in a research for creating an annotated chat corpus with both lexical and semantic tags to facilitate application of data mining in this domain. As another linguistic analysis example, Bogdanova et al. (2012a) have worked on identifying fixated discourse on chat logs. Fixated discourse associates to the strong intention of the predator to keep the focus of the conversation on sexual topics. Finally, a holistic approach has been presented by Cano et al. (2014) based on leveraging lexical features, sentiment features, content and psycho-linguistic features and discourse patterns. They have used semantic frames, which incorporate the general aspects of a discourse, and added them as additional features to the original bag-of-word model.

In the remainder of the chapter, we discuss the methods mentioned above in greater detail and highlight their strengths and weaknesses. Table 1 shows a mapping between data mining techniques and their applications in predator identification problem as well as the reference to corresponding works which have been done in the domain.

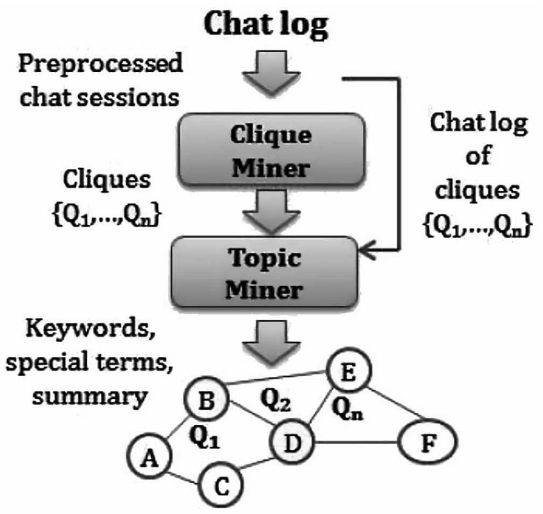
2.1 Criminal Network Analysis and Visualization

Another aspect of OPI deals with pedophile covert network analysis and visualization. Iqbal et al. (2012) have used the concept of criminal clique mining on chat logs to reveal the hidden relationship among criminals. Figure 4 depicts the main components of their framework.

Table 1. Mapping of applications in OPI and corresponding data mining techniques

Application Area in OPI	Data Mining Technique(s)	Previous Works
Predator Detection	Statistical Binary classification	• Approaches in PAN-2012 • (Morris, 2013) • (Cano et al., 2014) • (Pendar, 2007)
	Latent Semantic Indexing	• (Kontostathis et al, 2013)
	Rule-based approach	• (Mcghee et al., 2011)
• Criminal Network Analysis • Criminal Network Visualization	Graph Mining	• (Iqbal et al., 2012) • (Carley, 2015)

Figure 4. Overview of the framework for mining criminal networks in chat logs
Iqbal et al., 2012.



To analyze a criminal social network we need to explore the communication structure and the patterns by which network communications evolve. Based on the work of Klerks (2003), criminal network investigation approaches can be categorized into three types:

1. Manual approach,
2. Graphic-based approach, and
3. Social network analysis.

Since the focus of this chapter is on the identification aspects rather than network visualization and analysis, we focus on the predator detection through the remaining sections.

2.2 Successful Sample Tools

In order to introduce successful examples of automated online predator identification we briefly describe two software tools in which data mining techniques have been applied to this domain in a practical environment.

- **ChatCoder 2:** This tool was implemented at Mathematics and Computer Science Department of Ursinus College in Pennsylvania in 2011(Mcghee et al., 2011). The software leveraged a rule-based approach to classify the messages in forums into several categories including 'exchange of personal information', 'grooming', 'approach', and 'none of the categories'. These categories have been chosen based on the different phases defined in communication theory described in section 1-2-2. This software system could provide 68% accuracy on a public dataset available by a non-profit organization called perverted justice. One of the success factors of the software is that the system outperformed the human labeling in some of the categories.
- **Child Exploitation Tracking System (CETS):** In 2003, Microsoft initiated a new tool called Child Exploitation Tracking System (CETS) through close collaboration with Toronto Police Service to create an infrastructure for sharing the relevant documents and evidences among different investigators. According to the Royal Canadian Mounted Police (RCMP) the tool is still being extensively used in Canada as a cross-jurisdiction information sharing system between child exploitation investigators (Toews, 2013). CETS can be construed as a reliable tool that is being effectively utilized for combating child exploitation.

In the remainder of this chapter, we will describe the data mining methods used for preprocessing textual data and common algorithms for predator identification.

3. PREPROCESSING METHODS FOR CHAT LOGS

All data mining processes require a domain specific data preprocessing task which is often a tedious and time consuming task. Preprocessing of chat logs includes different subtasks ranging from parsing the raw textual log files to removing noise and reducing dimensionality.

3.1 Chat log's Data Format

Since analyzing textual data is less expensive and more efficient than analyzing other media such as image and video, currently almost all of the approaches for OPI use solely textual data. The logs format strongly depends on the software tool which is used for logging the conversations on the chat server. Although it can be any log format depending on the logging tool, usually the raw chat log data is gathered in a semi-structured textual format such as XML or Jason file. Various items can be stored in these logs among which the following three elements are essential for OPI analysis:

- **Authors:** Participants who are usually identified by unique identifiers.
- **Message Text:** The textual transferred message.
- **Time Stamps:** The date and time corresponding to each exchanged message.

Figure 5 shows a sample template for a chat log in XML format.

3.2 Noise Removal

Noise removal procedure is done for the purpose of improving the performance of learning (in terms of precision, recall, and accuracy) and also reducing the training time. In OPI problems, the noise removal procedure falls into one of the following categories or a combination of both:

Figure 5. A simple chat log in XML format with essential items for OPI

```
<Log date=May 05,2015>
    <message id=1>
            <participant id= Q12445678D></participant>
            <text> hi </text>
            <time>May 05,2015-16:45:20</time>
    </message>
    <message id=2>
            <participant id= Q12445678D></participant>
            <text> r u there? </text>
            <time>May 05,2015-16:45:26</time>
    </message>
    <message id=3>
            <participant id= F11445E211></participant>
            <text> hi </text>
            <time>May 05,2015-16:45:20</time>
    </message>
</Log>
```

1. **Removing Noisy Conversations:** This category of noise removal procedure includes identifying and eliminating useless samples that do not affect the learning process. This includes removing the following items:
   a. **Non-Textual Samples:** Real-world chat logs may contain conversations with only non-textual data or a very tiny amount of textual information. These samples can be safely ignored.
   b. **Conversations which Include Only One Participant:** This kind of conversations usually exists in a chat log corpus due to the fact that a participant may be unsuccessful in having a conversation with another participant.
   c. **Extremely Short Messages:** E.g. those which only contain a normal short greeting between two or more participants.
2. **Removing Noisy Features:** This category includes removal procedures for eliminating noise from features obtained during the feature extraction procedure (see section 4). Feature extraction and its corresponding methods in OPI will be discussed in the next section. For now, features can be considered as the set of important terms in a conversation. Noisy features may include the followings:
   a. Terms which are not in the proper encoding (This happens especially when there are multiple languages involved in the training corpus or there are other encodings than Unicode Transformation Format (UTF).)
   b. Small images or emoticons transferred among a whole bunch of textual conversations are construed as noisy features depending on the approach. It should be noted that emoticons are considered as a valuable source of information especially for extracting sentiment features. In such cases, emoticons should not be treated as noise.
   c. Unintentional Misspelled words throughout the conversation. It is worth mentioning that *intentional* misspelled words often play an important role in this application domain. As Villatoro-Tello Villatoro-Tello et al. (2012) state: For example in the grooming phase the perpetrator may amend the relationship by an emphasized "soryyyyyyyyy" when the child felt threatening by any obtrusive language *(p. 4). On the other hand, differentiation of intentional from unintentional spelling errors is not an easy task. Therefore, some researchers avoid the entire spell checking in the hope of gaining quality improvement.*

3.3 Feature Selection and Dimensionality Reduction

Feature selection involves a procedure to select a smaller subset of terms from a large set of candidate terms. The process is applied in order to increase the accuracy and/or decrease the training time (see section 5).

Let *D* be a data set with set of *n* terms denoted by *N*, and also let *f* be the function which maps the conversations into $l \in \{p, np\}$ in which *P* and *NP* represent the predatory instances and non-predatory instances respectively. Feature selection can be defined as the process of finding $N^* \subseteq N$ so that the performance of classifier *f* is maximized. Performance of classifier is typically defined by accuracy, precision, recall, and f-measure. Forman describes a holistic introduction of feature selection techniques used in text classification (Forman, 2003). Two common feature selection techniques are widely used in OPI domain. One of them falls into the category of supervised feature selection techniques and the other one is an unsupervised technique:

- **Unsupervised Feature Selection:**
  - **Document Frequency Thresholding:** Let *d*(*t*) be the number of documents in which term *t* occurs. Subset N^* contains ***t*** if and only if $d(t) \geq l$, in which $l \in \mathbb{N}$ is an arbitrary threshold. As Yang and Pedersen (1997) state, Document frequency thresholding is the simplest technique which scales well to large corpora.
- **Supervised Feature Selection:**
  - **Information Gain:** Supervised feature selection techniques measure each feature based on its contribution to identification of the correct category (i.e. predatory or non-predatory). Among numerous supervised techniques for feature selection in text classification task, information gain can be considered as the most successful technique in the domain of OPI. The notation used by Forman (2003) for calculating information gain in binary classification problem suits well in OPI. According to this notation, information gain for a specific term (feature) is defined as:

$$IG(t) = e(pos, neg) - \left[P(t)e(tp, fp) + P'(t)e(fn, tn)\right]$$

where:

$$e(x, y) = -\frac{x}{x+y}\log_2\left(\frac{x}{x+y}\right) - \frac{y}{x+y}\log_2\left(\frac{y}{x+y}\right)$$

In addition, *pos* and *neg* represent the number of predatory and non-predatory cases respectively. Also *tp* represents the number of predatory cases containing the term, *fp* is the number of non-predatory cases containing the term, *fn* is the number of predatory cases not containing the term and *tn* represents the number of negative cases not containing the term. $P\left(t\right)$ is calculated as follows:

$$p\left(t\right) = \frac{no.\ of\ cases\ containing\ term\ t}{total\ no.\ of\ cases}; p'\left(t\right) = 1 - p\left(t\right)$$

There is a large variety of dimensionality reduction techniques used in data mining but there is a text-specific dimensionality reduction technique called stemming which is widely used in text mining. The purpose of doing this preprocessing step is to reduce the terms 'work', 'works', 'worker' and 'working' into one dimension as 'work'. This process usually has a desirable effect on the performance of text categorization both in terms of quality and time efficiency. However, unfortunately this technique may not provide the same desirable effect for the OPI problem domain due to the fact that it will distort the information pertaining to the writing style of predators in chat logs. Accordingly, the best results have been reported while stemming has been avoided.

4. FEATURE EXTRACTION

Table 2 categorizes the features which are used in mining OPI problems along with corresponding previous works that have utilized these features.

In the remaining part of this section, the above categories are dissected in greater detail.

4.1 Lexical Features

Lexical features are word-related features that are directly extracted from the sentence. These words are used as candidate features in classification algorithms in order to determine the category to which chat logs belong. One of the simplest approaches used in OPI problem to extract features is the *bag-of-words approach* that treats a chat log as a set of words. Each word that exists in a chat log is considered as a candidate feature and then these features are weighted in terms of their frequency of occurrence. Typically, the initial candidate feature set is built by extracting *n*-grams (unigrams, bigrams, and trigrams) from the training data. Usually words that have insubstantial lexical meaning or known as stop words are filtered out during pre-

Table 2. Categorization of features which are used in OPI problem

Feature Category	Description	Application Examples
Lexical Features	Bag-of-word representations including: • Unigrams • Bigrams • Trigrams and other N-grams	• (Villatoro-Tello et al., 2012) • (Morris, 2013) • (Pendar, 2007)
Behavioral Features	• The number of times this author initiates a conversation • The number of times the author asks a question • Response Time • Conversation Dominance • Number of turn-takings	• (Morris, 2013)
Psychological and Linguistic Features	• Fixated Discourse (see below) • Writing Style (see below) • Emoticons (see below) • Tendency to change conversation to sexual discourse • Awareness of doing an illegal and non-moral action that may cause prosecution • Mimicking children language	• (Bogdanova et al., 2012a) • (Mcghee et al., 2011) • (Hogenboom et al., 2013)
Sentiment-oriented Features	Fear, Anger, Anticipation, Joy, Sadness, Disgust, Surprise, etc.	• (Bogdanova et al., 2012b)

processing in standard text categorization and Information Retrieval (IR) studies, but because of the fact that chat logs have their own vocabulary and rules, which makes general stop word list non-functional, researchers usually create their own list of stop words. As an example, the stop word list used by Pendar (2007) contained specifically 79 most frequent words in the corpus. After filtering out the entire stop words successfully, the *n*-grams and their corresponding frequencies for each chat log (or for each chat participant) are extracted.

Standard bag-of-words model has been shown to be robust in a wide variety of text classification problems. *Term Frequency-Inverse Document Frequency (TF-IDF),* is the most common weighting approach that is extensively used for weighting the candidate features before performing any feature selection procedure. In this weighting scheme, the most important words tend to have higher weights. This is implicitly achieved by multiplying the frequency of term *t* by a magnitude which is inversely proportional to the occurrence of term *t* in the whole corpus. Generally speaking, building unigrams and bigrams (pairs of consecutive words) produce better results than higher *n*-grams. Depending on the training corpus, the use of bigrams may increase the performance at the expense of increasing the size of the feature-space. However, for a training problem it may turn out that unigram model produces a better performance. Regardless of the classification performance, the size of the feature set for bigram representation is typically much bigger than that of unigram model. The resultant feature set can also be enriched by adding *domain-specific* features.

As an instance, in the work of Morris (2013), special tokens such as \SMILEY, \MALE.name, \FEMALE.name, \NUM and \PHONE.name were added to the lexical features in order to enrich the initial feature set. However, it was mentioned that these enhancements seemed to add a non-remarkable improvement.

Considering the inadequacy of bag-of-words models in reflecting deep semantic notions hidden in the conversations, one can also enrich the feature set with behavioral features which would be explained in detail below.

4.2 Behavioral Features

In this part, we list the high-level behavioral features and their applicability in the detection of online predators. Behavioral features are characterized as features that capture the actions of a user within a conversation. Morris (2013) has classified behavioral features as: 'Initiative', 'Attentiveness' and 'Conversation dominance' for which the details are given accordingly:

- **Initiative:** This can be measured by *number of initiations* (i.e. number of times a specified participant starts the conversation), *initiation rate* (i.e. the ratio of number of initiations to the whole number of conversations), questions and question rate in order to understand the author's tendency during the conversation.
- **Attentiveness:** This feature corresponds to the mean, median, and max response times for each author.
- **Conversation Dominance:** A set of features such as 'Message Ratio', 'Word Count Ratio' that reflect the degree to which the focal author dominates the conversation.

In order to successfully distinguish predators from victims, the above mentioned features are critical for *'symmetry-breaking'* (Morris, 2013). That is, given the fact that two authors in a chat conversation use very similar languages, behavioral features are one of the significant identifiers or non-lexical aspects of the conversation which are able to differentiate predators from victims.

4.3 Psychological and Linguistic Features

Psycho-linguistic features form another important aspect of feature extraction in OPI domain. 'Fixated discourse' is one of the most prominent psychological features used in OPI. Bogdanova et al. (2012a) defined fixated discourse as the unwilling-

ness of the predator to change the topic. For instance, predators always ignore the question or interruptions of pseudo-victim and has the tendency to go back to the sex-related conversation. According to them, chat logs might include implicit and explicit sexual content as predators gradually alter the direction of conversation to sex by starting with some ordinary compliments. On the other hand, predators often are aware of the fact that what they do on chat rooms is not moral and they try to transfer the responsibility of the victim and they often behave as children by copying the children's linguistic style (Bogdanova et al., 2014).

The following analysis of chat logs identifies the important characteristics of predators' psychological and linguistic features (Bogdanova et al., 2014):

- ***Implicit/Explicit Content:*** *Typically, pedophiles shift gradually to the sexual conversation, starting with ordinary compliments and then they shift the conversation to make it overtly related to sex. They do not hide their intentions.*
- ***Fixated Discourse:*** *Pedophiles are reluctant to step aside from the sexual conversation. In other words, pedophiles try to come back to the sex-related conversation when the victim steps outside of the topic.*
- *Offenders often understand that what they are doing is not moral.*
- *They transfer responsibility to the victim.*
- *Offenders often behave as children, copying their linguistic style. Colloquialisms appear often in their messages.*

Mcghee et al. (2011) have proposed the following linguistic features which were denoted as 'Writing Style' in Table 2: Total number of words in a line, number of first-person pronouns, second-person pronouns or third-person pronouns, number of personal information nouns *(e.g., age, pic)*, number of relationship nouns *(e.g., boyfriend, date)*, number of family nouns *(e.g., mom, sibling)*, number of communicative desensitization words *(e.g., kiss, bra)*, number of approach verbs *(e.g., meet, see, hotel)*.

Hogenboom et al. (2013) mentioned that their research had shown that people were influenced by the nonverbal cues and emoticons. These are widely used to express sentiments such as happiness, sadness, joy or anger, therefore emoticons could also be used to reveal the predators' sentiments and their tendencies in order to be dominant in the conversation and also to copy children's' behavior as explained above. Table 3 shows the typical examples of emoticons and their sentimental interpretations (Hogenboom et al., 2013).

Table 3. Typical examples of emoticon synsets

Emoticon Synset	Emoticons
Happiness	:-D, =D, xD, (^_^)
Sadness	:-(, =(
Crying	:'(, ='(, (;_;)
Boredom	-_-, -.-, (>_<)
Love	<3, (L)
Embarrassment	:-$, =$, >///<

Hogenboom et al., 2013.

4.4 Sentiment-Oriented Features

In addition to the emoticons explained earlier, sentiment of chat logs can provide significant markers in terms of predator identification and unveil other important semantic dimensions. According to Bogdanova et al, in general, predatory conversations contain more positive and less negative words. The sentiments and emotional markers shown in Table 4 were used as features in their experiments.

5. LEARNING PREDATORY PATTERNS

First, we formally introduce the notion of binary classification which is used in OPI problem. Let data set *D* be defined as $D \subset X \times Y$ where $X = \left\{X_1, X_2, ..., X_m\right\}$ is the set of *m* observation vectors $X_i; i \in \left\{1, 2, ..., m\right\}$, so that $X_i = \left(x_1, x_2, ..., x_n\right)^T$

Table 4. Sentiment features

Feature	Example
Positive words	Cute, pretty
Negative words	Dangerous, annoying
JOY words	Happy, cheer
SADNESS words	Bored, sad
ANGER words	Annoying, furious
SURPRISE words	Astonished, wonder
DISGUST words	Yucky, nausea
FEAR words	Scared, panic

Bogdanova et al., 2014.

is the corresponding vector of i^{th} observation containing n feature values $X_j; j \in \{1,...,n\}$. Also $Y = \{p, np\}$ is the set of two class labels corresponding to predatory and non-predatory instances respectively. The classification task is defined as finding a mapping function $f : X \to Y$ such that $f(X)$ is able to predict $y \in Y$ as accurately as possible.

As a typical approach, the data is split into training and testing sets and the classification model learns from the training set and then applies to the test set to evaluate the performance of classification. A wide variety of learning algorithms for learning function *f* have been proposed and utilized in data mining. In the following parts of this section, we introduce the most common algorithms that have been used for solving OPI problems.

5.1 OPI Standard Classification Methods

This part is dedicated to introducing concrete data mining classification algorithms which have been used in OPI problems. First, we describe a highly-standard probabilistic algorithm called Naïve Bayes which has been used for text classification since long time ago. Then we discuss the most intuitive algorithm which is usually used in information retrieval called K-NN. Then we proceed to more advanced algorithms such as Entropy-based classification, Support Vector Machines and Artificial Neural Networks.

5.1.1 Naive Bayes

This algorithm is used extensively as the baseline in text classification studies. This means that researchers accept it as an efficient algorithm and aim to improve its performance through other novel algorithms. Although the reader can refer to Duda et al. (2012) for a thorough explanation of the algorithm, we present a simple description here. Let *D* be a data set defined in the previous section. Assuming that all discrete-valued features are conditionally independent given the class label ***y*** (known as '*Naive Bayes assumption*') we can simplify $P(X \mid y)$ that is the conditional probability of observation vector *X* given *y* as follows:

$$P(X \mid y) = P(x_1, x_2, ..., x_n \mid y) = P(x_1 \mid y) P(x_2 \mid x_1, y) ... P(x_n \mid y, x_1, x_2, ..., x_{n-1})$$
$$= P(x_1 \mid y) P(x_2 \mid y) ... P(x_n \mid y) = \prod_i P(x_i \mid y); \textit{by Naive Bayes assumption}$$

Having $P\left(X \mid y\right)$ calculated as above, one can predict the most likely class label $\left(y^*\right)$ by using the Bayes rule as follows:

$$y^* = \arg\max_y P\left(y\right)P\left(X \mid y\right) = \arg\max_y P\left(y\right)\prod_i P\left(x_i \mid y\right)$$

Specifically in OPI problems, we deal with labels *p* and *np* and accordingly we can rewrite the above as:

$$y^* = \max\left(P\left(y = p\right)\right)\prod_i P\left(x_i \mid y = p\right), P\left(y = np\right)\prod_i P\left(x_i \mid y = np\right)$$

5.1.2 K-Nearest Neighbor

Kang et al. (2012) have used K-Nearest Neighbor (also known as KNN) for online predator identification. They have used a weighted modification of classic KNN. Although their result is not comparable with other methods, their approach is worth mentioning due to the rational justification behind it. The simplest version of the algorithm can be outlined as shown in Algorithm 1, where majority function simply calculates the majority of class labels among neighbors. A more sophisticated explanation can be found in Duda et al. (2012).

There are also weighted versions of KNN which assigns different weights to the neighbors based on its distance to the query point (i.e. the instance that might be whether predatory or non-predatory). The authors of Kang et al. (2012) state that choosing a good number of *k* which can provide good results both on training set and testing set still remains as a challenge.

5.1.3 Maximum Entropy Classification

The conditional independence assumption of Naive Bayes is not realistic at least in dealing with textual documents. Therefore, we may need statistical models that can consider the notion of dependent random variables. Maximum Entropy Classifier is a *discriminative* approach which tries to build a statistical model of conditional probability distribution $P\left(y \mid X\right)$. There is an infinite number of such models, but based on the maximum entropy principle, the best model is the one which maximizes entropy of $H\left(P\right)$. Berger et al. (1996) formally state this as follows:

$$H\left(P\right) \equiv -\sum_{x,y} \tilde{P}(x)P\left(y \mid X\right)\log P\left(y \mid X\right)$$

Algorithm 1. KNN (D, x, k, m)

```
    Input:
       D ← Training data set
        χ ← Query vector
       m ← Distance measure (e.g. Euclidean Distance)
    Output:
        y: predicted class label 'p' or 'np' (i.e. deciding
whether the sample χ is predatory or not)
begin
1.         N, L ← ∅
2.         Find k nearest neighbors to x based on the distance
measure m and add them to neighbors set N.
3.         Add the corresponding class labels of found neigh-
bors in N to Labels set L.
4.         y ← Majority (L)
end
```

In order to choose the best model P^* from set C (the set of valid probability distributions), based on the maximum entropy selection, we have to solve the following optimization problem:

$$P^* = \arg\max_{P \in C} H\left(P\right)$$

Note that in Berger's notation the probability distributions are denoted by *p* instead of P. However, we use *P* for denoting the distributions since we have already reserved symbol *p* for representing the 'predatory' class label.

Also note that the above optimization problem is a constrained one. Nevertheless, describing the complete theoretical background of Maximum Entropy Classifier is beyond the scope of this chapter. A keen reader may like to refer to Berger's et al. (1996) seminal paper about introducing max entropy on natural language processing.

Eriksson and Karlgren (2012), and Kern et al. (2012) used Maximum Entropy Classifier in the OPI domain on PAN-2012 data set. Their obtained results are comparable with the winner's approach of PAN-2012 and can be considered as a successful approach for solving OPI problems.

5.1.4 Support Vector Machines

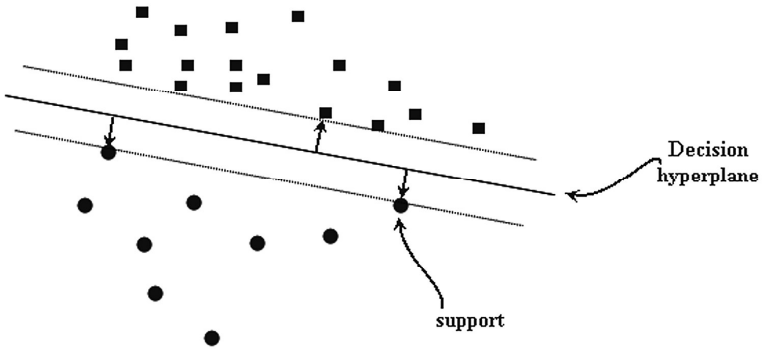
Currently, Support Vector Machines (SVMs) have shown the best results among all different classification algorithms which have been used for OPI problems. SVM was explicitly introduced by Vapnik, (1995) in his book entitled *'The Nature of Statistical Learning Theory'*. The goal of SVM algorithm is to obtain a hyperplane which maximizes the margins between positive and negative instances (predatory and non-predatory samples in our case). The margin is defined as two times the distance from the decision hyperplane. Figure 6 illustrates the idea of SVM.

Instances which are located on the margins are called 'support vectors'. The main result of training an SVM model is to obtain these support vectors also known as *supports*. These instances are the only ones that take into considerations when a new instance is supposed to be classified by the trained model. The training procedure contains solving a constrained optimization problem which is usually solved by out-of-the-shelf quadratic programming tools. SVM has been used by Morris (2013), and Villatoro-Tello et al. (2012) and as already mentioned, SVM has the highest performance among algorithms used in PAN-2012. The reader may refer to Bishop (2006) for the mathematical background of support vector machines. Note that besides the classification algorithm, preprocessing methods and feature extraction methods, described in previous part, have also drastic impact on the final performance.

5.1.5 Neural Networks

Even though there are a large variety of neural network algorithms, the one that has been successfully utilized in OPI is the Multi Layer Perceptron (MLP). In a pioneering work, the researchers tested MLP for identifying predatory conversations and also

Figure 6. A toy example showing the output of SVM for a binary classification as well as the margin and three support vectors. Squares and circles represent predatory and non-predatory samples respectively



identifying the likely predators and then compared the obtained performance with that of Support Vector Machines (Villatoro-Tello et al., 2012). The results showed that the performance of neural network and SVM are so close to each other. More specifically, the neural network outperformed SVM in identifying predators versus victims and on the contrary, SVM performed better in identifying the predatory conversations. Here, we introduce the key concepts of MLPs without a detailed description of algorithm.

Units are considered as building blocks of MLPs. Each unit resembles a neuron that has a specific 'activation function' which generates the final output of a neuron. In a higher level of abstraction, the network contains several layers of units including input, output and one or more hidden units. The units in a layer are not connected to each other, while usually all of the units in a previous layer are connected to the units in the next layer. There is a real number assigned to each connection called 'weights'. The final output z_j of a neuron located in the first hidden layer with M units is calculated as follows (Bishop, 2006):

$$z_j = h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right)$$

where h is the activation function for neurons in the hidden layer which is usually chosen to be a sigmoidal function. Also D is the number of features in the input sample. Finally, superscript (1) denotes the layer in which neuron is located. This process of linear transformations continues in a cascading manner from previous layers to the last hidden layer and eventually the output layer in which the final output of the network is generated.

The learning algorithm of an MLP with a specified structure is an algorithm which finds the relatively optimal network weights. The learning procedure encompasses a technique called 'backpropagation' and also an optimization technique called 'gradient descent'. For a complete description about Neural Networks, user could refer to Bishop's book (2006). Similar to SVM, there are out-of-the-shelf libraries and tools for building and training neural networks.

To summarize this section, we mention the outstanding algorithms among the above-mentioned methods. In terms of accuracy, precision, and recall rate, Support Vector Machines and Neural Networks (Multi-layer Perceptron) outperform other classification methods. However, it is suggested to always apply a simple approach such as Naïve Bayes to obtain a performance baseline so that if there is a problem with parameter tuning of the two mentioned algorithms, it will be revealed at the first stages of the analysis.

6. FUTURE RESEARCH DIRECTIONS

In spite of the considerable achievements mentioned in this chapter, there are still prominent challenges which researchers need to tackle in the field of OPI. Accordingly, we describe the potential future research lines based on our anticipation of the problem domain. In the following, we demonstrate the necessity for deeper linguistic analysis as well as the related challenges are discussed and then the newly emerged field of Web-based Dynamic Social Networks is introduced.

6.1 Performing Deeper Linguistic Analysis on Chat logs

Mining chat logs is strongly correlated to challenging problems in the domain of NLP including Word Sense Disambiguation (identifying the sense for a polysemic part of speech), Discourse Analysis (Discovering the conversation concepts and psychological characteristics of the writer), and also Named Entity Recognition (Extraction of role-playing entities such as locations, people and organizations).

On top of these linguistic challenges, there is another important issue related to the nature of chat logs: Conversational (i.e non-official) writing style of participants. Consider the following predatory conversation:

```
<text>i'm bored</text>
<text>Awww babe</text>
<text>I'm sorwy</text>
<text>where u at</text>
<text>Vegas</text>
<text>5-6 hours away</text>
<text>dude y cant u come then!?</text>
<text>I'm n vegas lol</text>
<text>I'm n another state</text>
<text>I'm not n california</text>
<text>i thought u wanted 2 come c me</text>
<text>I do</text>
<text>But how can I went I'm n another state</text>
<text>when do u leave?</text>
<text>Dis mornin</text>
<text>well i guess u aint really my bf then cuz u lied</text>
```

This writing style requires some additional considerations that make it different from normal text mining. A common issue which arises in such a context is the existence of non-grammatical sentences which makes the typical parsing algorithms

inefficient. For instance, 'But how can I went I'm n another state'. Another issue is regarding the use of drastically misspelled words such as in 'Dis mornin'. Even using the sophisticated spell checkers or stemmers on such a data as a preprocessing phase would not be so efficient. Having too many different forms of writing for a single word causes the problem which is known as 'curse of dimensionality' which makes the learning algorithms significantly inefficient.

Another important issue which might not be so related to the linguistic aspects of OPI is the imbalanced nature of chat logs data. This means usually there are too many non-predatory instances compared to predatory ones. This problem makes the learning process more challenging since it requires specific algorithms to deal with this type of imbalanced data.

6.2 Learning Deep Architectures

Recently there has been a hot trend among researchers in the field of artificial intelligence and machine learning about a new paradigm of learning which can mimic the behavior of human brain or human visual system in a more accurate way (Bengio, 2009; Schmidhuber, 2015). The new learning way is called deep learning because of the fact that there is a hierarchy of numerous layers in the main model and each layer encodes a level of abstraction in the training data. Using these models has been proven to be more efficient than the simple data mining and machine learning models mentioned above. Accordingly, we anticipate that these models will be utilized in the field of OPI in the near future.

6.3 Web-Based Dynamic Social Networks

Criminal social network analysis and visualization was briefly mentioned previously. Unlike the traditional criminal networks which have a strictly hierarchical structure, online pedophile networks naturally have a cellular and distributed structure and usually do not have obvious leaders. These special types of networks demand the usage of approaches specifically designed for tackling with the cellular distributed crime networks. These approaches should be able to analyze smaller crime networks that do not necessarily have a specific powerful leader. A new tool for analyzing this sort of networks has been developed by Carley (2015) at Carnegie Mellon University which might be useful for analyzing pedophile covert networks. In addition to the approaches identified by Klerks in section 1, a new branch of social network analysis called Web-based Dynamic Social Network has been revealed recently to address the mentioned requirement. In this point of view, WDSN differs from traditional social networks in the sense that they are cellular, distributed, web-based, dynamic, and may contain varying levels of uncertainty. According to Berger-Wolf and Saia

(2006), dynamic network analysis enables probabilistic reasoning about changes in dynamic networked web-based communities and how such networks evolve, adapt to changes, and how they can be destabilized. Leveraging WDSN to identify pedophile covert networks and analyze their evolving network communication structure can be considered as one of the most significant directions in the field.

According to the above-mentioned challenges, we anticipate the future of this field spin around the following issues:

- Achieving a deeper understanding of text, or more generally natural language, to uncover the semantics behind the chat logs and improve the accuracy of classification models.
- Leveraging deep learning as a new trend in artificial intelligence for building more sophisticated language models from chat logs.
- Using the concepts of WDSN introduced above to identify pedophile covert networks and analyze their evolving network communication structure.

7. CONCLUSION

Rapidly growing penetration of online communications in juveniles' daily lives makes it vital to leverage data mining techniques for automatic identification of online predators. Automated investigation of chat logs is one of the most proactive and effective approaches that can be used to avoid the consequences of related crimes. The most popular preprocessing techniques including noise removal, feature selection, and dimensionality reduction were introduced. Also, different aspects of suitable feature extraction procedure for this problem domain were discussed and finally the most common data mining classification algorithms which are frequently used in OPI were introduced. We predict that the future of this research line in the next decade will spin around Social Network Analysis featured by deeper linguistic analysis to understand the semantics of messages.

ACKNOWLEDGMENT

This research was supported by grants from the Natural Science and Engineering Research Council of Canada and Concordia University.

REFERENCES

Aggarwal, C. C. (2015). Mining Text Data. In *Data Mining* (pp. 429–455). Springer International Publishing.

Bengio, Y. (2009). Learning Deep Architectures for AI. *Found. Trends Mach. Learn.*, *2*(1), 1–127. doi:10.1561/2200000006

Berger, A. L., Pietra, V. J. D., & Pietra, S. A. D. (1996). A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*, *22*(1), 39–71.

Berger-Wolf, T. Y., & Saia, J. (2006). A Framework for Analysis of Dynamic Social Networks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 523–528). New York, NY: ACM. doi:10.1145/1150402.1150462

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Bogdanova, D., Rosso, P., & Solorio, T. (2012a). Modelling Fixated Discourse in Chats with Cyberpedophiles. In *Proceedings of the Workshop on Computational Approaches to Deception Detection* (pp. 86–90). Stroudsburg, PA: Association for Computational Linguistics.

Bogdanova, D., Rosso, P., & Solorio, T. (2012b). On the Impact of Sentiment and Emotion Based Features in Detecting Online Sexual Predators. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis* (pp. 110–118). Stroudsburg, PA: Association for Computational Linguistics.

Bogdanova, D., Rosso, P., & Solorio, T. (2014). Exploring high-level features for detecting cyberpedophilia. *Computer Speech & Language*, *28*(1), 108–120. doi:10.1016/j.csl.2013.04.007

Cano, A., Fernandez, M., & Alani, H. (2014). Detecting Child Grooming Behaviour Patterns on Social Media. In L. Aiello & D. McFarland (Eds.), *Social Informatics* (Vol. 8851, pp. 412–427). Springer International Publishing. doi:10.1007/978-3-319-13734-6_30

Carley, K. M. (2015, May 8). *DyNet*. Retrieved from http://www.casos.cs.cmu.edu/projects/DyNet/dynet_info.html

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern Classification*. Wiley-Interscience.

Eriksson, G., & Karlgren, J. (2012). *Features for modelling characteristics of conversations*. Presented at the Notebook for PAN at CLEF 2012, Rome, Italy.

Escalante, H. J., Villatoro-Tello, E., Juárez, A., Montes-y-Gómez, M., & Villaseñor, L. (2013). Sexual predator detection in chats with chained classifiers. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 46–54). Atlanta, GA: Association for Computational Linguistics.

Forman, G. (2003). An Extensive Empirical Study of Feature Selection Metrics for Text Classification. *Journal of Machine Learning Research*, *3*, 1289–1305.

Forsyth, E. N., & Martell, C. H. (2007). Lexical and Discourse Analysis of Online Chat Dialog. In *Semantic Computing, 2007. ICSC 2007. International Conference on* (pp. 19–26).

Hogenboom, A., Bal, D., Frasincar, F., Bal, M., de Jong, F., & Kaymak, U. (2013). Exploiting Emoticons in Sentiment Analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing* (pp. 703–710). New York, NY: ACM. doi:10.1145/2480362.2480498

Inches, G., & Crestani, F. (2012). *Overview of the International Sexual Predator Identification Competition at PAN-2012*. CLEF (working notes).

Iqbal, F., Fung, B. C. M., & Debbabi, M. (2012). Mining Criminal Networks from Chat Log. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on* (Vol. 1, pp. 332–337). doi:10.1109/WI-IAT.2012.68

Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A., … Li, H. (2015). A survey on text mining in social networks. *The Knowledge Engineering Review, 30*(2), 157–170.

Kang, I.-S., Kim, C.-K., Kang, S.-J., & Na, S.-H. (2012). *IR-based k-Nearest Neighbor Approach for Identifying Abnormal Chat Users.* Presented at the Notebook for PAN at CLEF 2012, Rome, Italy.

Kern, R., Klampfl, S., & Zechner, M. (2012). *Vote/Veto Classification, Ensemble Clustering and Sequence Classification for Author Identification.* Presented at the Notebook for PAN at CLEF 2012, Rome, Italy.

Kierkegaard, S. (2008). Cybering, online grooming and ageplay. *Computer Law & Security Report*, *24*(1), 41–55. doi:10.1016/j.clsr.2007.11.004

Klerks, P. (2003). The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands. In *Transnational Organised Crime Perspectives on Global Security*. London: Routledge.

Kontostathis, A. (2009). Toward the tracking and categorization of internet predators. In *Proceeding of Text Mining Workshop 2009 held in conjunction with Ninth Siam International Conference Data Mining*.

Kontostathis, A., Reynolds, K., Garron, A., & Edwards, L. (2013). Detecting Cyberbullying: Query Terms and Techniques. In *Proceedings of the 5th Annual ACM Web Science Conference* (pp. 195–204). New York, NY: ACM. doi:10.1145/2464464.2464499

Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 415–463). Springer, US. doi:10.1007/978-1-4614-3223-4_13

Mcghee, I., Bayzick, J., Kontostathis, A., Edwards, L., Mcbride, A., & Jakubowski, E. (2011). Learning to Identify Internet Sexual Predation. *International Journal of Electronic Commerce*, *15*(3), 103–122. doi:10.2753/JEC1086-4415150305

Michalopoulos, D., & Mavridis, I. (2011). Utilizing document classification for grooming attack recognition. In *Computers and Communications (ISCC), 2011 IEEE Symposium on* (pp. 864–869). doi:10.1109/ISCC.2011.5983950

Morris, C. (2013, January 30). *Identifying Online Sexual Predators by SVM Classification with Lexical and Behavioral Features* (Master of Science Thesis). University of Toronto, Canada. Retrieved from ftp.cs.toronto.edu/pub/gh/Morris,Colin-MSc-thesis-2013.pdf

Olson, L. N., Daggs, J. L., Ellevold, B. L., & Rogers, T. K. K. (2007). Entrapping the Innocent: Toward a Theory of Child Sexual Predators' Luring Communication. *Communication Theory*, *17*(3), 231–251. doi:10.1111/j.1468-2885.2007.00294.x

PAN-2012. (n.d.). Retrieved from http://pan.webis.de

Pendar, N. (2007). *Toward spotting the pedophile telling victim from predator in text chats*. Washington, DC: IEEE. doi:10.1109/ICSC.2007.32

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. doi:10.1016/j.neunet.2014.09.003 PMID:25462637

Toews, V. (2013). *Royal Canadian Mounted Police Report on Plans and Priorities 2013-2014*. Canada: RCMP. Retrieved from http://www.rcmp-grc.gc.ca/rpp/2013-2014/rpp-eng.htm

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York, NY: Springer-Verlag New York, Inc. doi:10.1007/978-1-4757-2440-0

Villatoro-Tello, E., Juárez-González, A., Escalante, H. J., Montes-y-Gómez, M., & Villaseñor-Pineda, L. (2012). *A Two-step Approach for Effective detection of Misbehaving Users in Chats*. Presented at the Notebook for PAN at CLEF'12, Rome, Italy.

Yang, Y., & Pedersen, J. O. (1997). A Comparative Study on Feature Selection in Text Categorization. In *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 412–420). San Francisco, CA: Morgan Kaufmann Publishers Inc.

KEY TERMS AND DEFINITIONS

Age Disparity: The significant difference between the age of predator adult and that of minor (i.e. victim).

Conversation: A chat session that encompasses the messages exchanged between participants.

Clique: In the context of graph theory, a clique is a set of vertices which their corresponding subgraph is complete (i.e. fully connected). While, in the context of mining criminal networks this definition can be manipulated to less mathematical definitions. Specifically, in the context of chat logs, a clique might be defined as a set of persons who participate in a minimum number of chat sessions.

Luring Communication Theory: A communication theory that models the behavior of predator for approaching, entrapping, and establishing the predatory relationship with minor.

Maximum Entropy Principle: The principle that states the best probability distribution for a statistical model is the one that has the maximum entropy.

Minor: A Person under the age of 18 who is considered as the potential victim of predatory attacks in cyber space.

Vector Space Model: The algebraic representation of documents based on their terms and the frequency of occurrence of each term.