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

Mohammadreza Keyvanpour
Alzahra University, Iran

Mohammadreza Ebrahimi
Concordia University, Canada

Necmiye Genc Nayebi
École de Technologie Supérieure (ÉTS), 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
Current Issues and Future Analysis in Text Mining for Information Security Applications
www.igi-global.com/chapter/current-issues-future-analysis-text/21752?camid=4v1a

Cost-Sensitive Learning in Medicine
Alberto Freitas, Pavel Brazdil and Altamiro Costa-Pereira (2009). *Data Mining and Medical Knowledge Management: Cases and Applications* (pp. 57-75).
www.igi-global.com/chapter/cost-sensitive-learning-medicine/7527?camid=4v1a