Expert Systems With Applications 89 (2017) 99–111

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# Early detection of deception and aggressiveness using profile-based representations

Hugo Jair Escalante[a,*], Esaú Villatoro-Tello[b], Sara E. Garza[c], A. Pastor López-Monroy[d], Manuel Montes-y-Gómez[a], Luis Villaseñor-Pineda[a]

[a] Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Luis Enrique Erro 1, Puebla 72840, Mexico
[b] Language and Reasoning Research Group, Information Technologies Department, Universidad Autónoma Metropolitana, Unidad Cuajimalpa (UAM-C), Cuidad de Mexico, 05348, Mexico
[c] School of Mechanical and Electrical Engineering (FIME), Universidad Autónoma de Nuevo León, San Nicolás de los Garza 66451, NL, Mexico
[d] Research in Text Understanding and Analysis of Language Lab, University of Houston, 4800 Calhoun Road, Houston, TX 77004, USA

ARTICLE INFO

*Article history:*
Received 10 October 2016
Revised 24 July 2017
Accepted 25 July 2017
Available online 25 July 2017

*MSC:*
68T10
68T20

*Keywords:*
Deception detection
Profile-based representations
Sexual predator detection
Aggressive text identification

ABSTRACT

*E*-communication represents a major threat to users who are exposed to a number of risks and potential attacks. Detecting these risks with as much anticipation as possible is crucial for prevention. However, much research so far has focused on forensic tools that can be applied only when an attack has been performed. This paper proposes a novel and effective methodology for the early detection of threats in written social media. The goal is to recognize a potential attack before it is consummated, and using a minimum amount of information. The proposed approach considers the use of profile-based representations (PBRs) for this goal. PBRs have multiple benefits, including non-sparsity, low dimensionality, and a proved discriminative power. Moreover, representations for partial documents can be derived naturally with PBRs, which makes them suitable for the addressed problem. Results include empirical evidence on the usefulness of PBRs in the early recognition setting for two tasks in which anticipation is critical: *sexual predator detection* and *aggressive text identification*. These results reveal, on the one hand, that PBRs achieve state of the art performance when using full-length documents (i.e., the classical task), and, on the other hand, that the proposed methodology outperforms previous work on early recognition of sexual predators by a considerable margin, while obtaining state of the art performance in aggressive text identification. To the best of our knowledge, these are the best results reported on early recognition for the approached problems. We foresee this work will pave the way for the development of novel methodologies for the problem and will motivate further research from the intelligent systems and text mining communities.



## 1. Introduction

Social media is perhaps the most used communication channel nowadays: anyone can express their opinion about any topic in any context (Kuz, Falco, & Giandini, 2016). In spite of this easiness of communication, this kind of media and – in general – e-communication media comprise a major threat to users, who are exposed to a number of risks and potential attacks. Consider, for example, the problem of detecting sexual predators approaching minors or the identification of aggressive users. These threats pose a challenge to the research community, that has to develop protective and preventive tools for avoiding potential risks.

A considerable amount of research has been devoted to detect these threats. However, current solutions work in a forensics scenario, i.e., they are applied once the attack has been accomplished. Although these solutions can be useful in certain contexts, preventive mechanisms would have a greater and immediate impact into user security.

Taking into account the latter scenario, this paper proposes a novel and effective methodology to detect potential attacks as early as possible (while communication is being performed). A difficulty that arises with early recognition tasks concerns information scarcity, since only partial information is available to detect the attack before it is consummated. To face this problem, the proposed approach considers the use of profile and subprofile-based representations. Under these representations, each term (e.g., word) is

* Corresponding author.
*E-mail addresses:* hugojair@inaoep.mx, hugo.jair@gmail.com (H.J. Escalante), evillatoro@correo.cua.uam.mx (E. Villatoro-Tello), sara.garzavl@uanl.edu.mx (S.E. Garza), alopezmonroy@uh.edu (A.P. López-Monroy), mmontesg@inaoep.mx (M. Montes-y-Gómez), villasen@inaoep.mx (L. Villaseñor-Pineda).

http://dx.doi.org/10.1016/j.eswa.2017.07.040
0957-4174/

associated to a vector that accounts for its semantics, where a document can be represented by aggregating the vectors of the terms it contains. As a result, documents and terms can lie in the same semantic space. Even when only a few terms are available, these representations can still be obtained – a convenient property that makes them suitable for early text classification. These representations, in addition, have the advantage of being non-sparse, low dimensional, and highly discriminative. This paper shows the benefits of using these representations to recognize the category of a document before it is available entirely. Specifically, the problems of *sexual predator* and *aggressive text* early recognition are approached. An extensive experimental evaluation reveals that the proposed methodology is able to obtain state of the art performance in the aforementioned tasks, while requiring a minimum amount of information from documents to make a decision. We foresee this work will pave the way for the development of novel methodologies for the problem, and will motivate further research from the intelligent systems and text mining communities.

The contributions of this paper are as follows:

- The use and performance evaluation of profile and subprofile-based representations for the problems of sexual predator detection and aggressive text identification. This is the first time that the previously mentioned representations are employed for these problems using full documents. It has been shown that state of the art performance can be obtained in the considered data sets, with the additional benefits of working with low dimensional and non-sparse representations.
- The use, adaptation, and suitability evaluation of profile and subprofile-based representations for the problem of *early text classification*. It is shown how they can naturally be used to represent documents containing partial information. This is the first time this feature is noticed and exploited. More importantly, results on early recognition performance outperform existing work in the sexual predator detection task by a large margin, while achieving comparable performance in the aggressive text detection problem.
- A comprehensive and extensive literature review on the automated detection of sexual predators and aggressive text in digital documents.

The rest of this paper is organized as follows. The next section provides a review of related work on automated deception detection in social media and early text classification. Section 3 describes the profile-based representations and how they are used for early recognition. Section 4 describes the experimental settings and the evaluation protocol. Section 5 reports experimental results and their analysis. Finally, Section 6 summarizes our main findings and outlines future work directions.

## 2. Related work

With the continued growth and use of Internet as a tool for communication worldwide, more and more people are enjoying and becoming more dependent on the convenience of its provided services. Unfortunately, the wide use of computers and mobile devices in conjunction with Internet has also been convenient to cyber-attackers. Nowadays, there are many types of attacks that an Internet user has to face: computer viruses, flaws in the operating system (backdoors opened), phishing, fraud activities, harassment, sexual predation, and other types of cyber-crimes. The common factor for all these activities is the easiness that attackers have to hide their real identities.

In this scenario, two problems are particularly relevant: the *identification of sexual predators* and the *detection of aggressive text*. Both of these are increasingly important issues especially because the target subjects are, commonly, under-age victims. Consequently, these problems are of great social concern and are becoming very important research problems, notably from the automatic *early* detection point of view. The remainder of this section provides a review of related work on both of the approached problems and the early recognition setting.

### 2.1. Sexual predator detection

Sexual predator identification is a critical problem given that the majority of cases of sexual assaulted children have agreed voluntarily to meet with their abuser (Finkelhor, Mitchell, Wolak, & Children, 2006). Clearly, the early detection of a malicious predatory behavior against a child could reduce the number of abused children. Traditionally, a term that is used to describe malicious actions with a potential aim of sexual exploitation or emotional connection with a child is referred to as "*Child Grooming*" or "*Grooming Attack*" (Kucukyilmaz, Cambazoglu, Aykanat, & Can, 2008). This attack is defined by Harms (2007) as "*a communication process by which a perpetrator applies affinity seeking strategies, while simultaneously engaging in sexual desensitization and information acquisition about targeted victims in order to develop relationships that result in need fulfillment*" (e.g. physical sexual molestation).

Given the difficulties involved in having access to useful data, i.e., where real pedophiles are involved, nowadays the problem of sexual predator identification by means of pattern recognition techniques is a relatively new research area. The usual approach to catch sexual predators is through police officers or volunteers who behave as fake children in chat rooms and provoke sexual offenders to approach them.[1] Unfortunately, online sexual predators always outnumber the law enforcement officers and volunteers. Therefore, tools that can automatically detect sexual predators in chat conversations (or at least serve as a support tool for officers) are highly needed.

Accordingly, there have been various proposed approaches to tackle this problem. These approaches fall into three main categories: *i)* identification of predatory chat lines, *ii)* classification of predatory chat conversations, and *iii)* identification of the offender and the victim. The following sections describe some of the most representative research works on each category.

#### 2.1.1. Identifying predatory chat lines

The proposed works under this category are based on the hypothesis that online child grooming by predators resembles the grooming stages used in face-to-face interactions (Black, Wollis, Woodworth, & Hancock, 2015). In general, three major stages are defined: *i) Gaining Access*: indicate predators intention to gain access to the victim; *ii) Deceptive Relationship*: indicate the deceptive relationship that the predator tries to establish with the minor, and are preliminary to a sexual exploitation attack; and *iii) Sexual Affair:* clearly indicate predator's intention for a sexual affair with the victim. Therefore, research works try to model the different grooming stages, and attempt to classify chat lines into some particular stage.

Research works that study the communicative strategies of online sexual predators are those from Kontostathis (2009) and Kontostathis et al. (2010), where the authors provide a tool that enables human annotators to label conversation lines into a particular grooming stage. Afterwards, a phrase-matching and a rule based approach is applied to classify a conversation as being associated (or not) to a grooming stage. Michalopoulos and Mavridis (2011) proposed a decision making method to be used for recognizing potential grooming threats by extracting information from

[1] The American foundation, called Perverted Justice (PJ) (http://www.perverted-justice.com/), follows the above mentioned approach.