Knowledge-Based Systems 96 (2016) 134–155

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A biology-inspired, data mining framework for extracting patterns in sexual cyberbullying data

N. Potha[a], M. Maragoudakis[a,*], D. Lyras[b]

[a] Department of Information and Communication Systems Engineering, University of the Aegean, Samos, 83200, Greece
[b] Faculty of Biology, Department II, Ludwig-Maximilians Universität, München Planegg-Martinsried, 82152, Germany

ARTICLE INFO

Article history:
Received 27 March 2015
Revised 23 December 2015
Accepted 27 December 2015
Available online 4 January 2016

Keywords:
Cyber bullying
Time series analysis
Hierarchical clustering
Symbolic representation

ABSTRACT

With the rapid growth of social media, users, especially adolescents, are spending significant amount of time on various social networking sites to connect with others, to share information, and to pursue common interests. However, as social networking has become widespread, certain people are finding illegal and unethical ways to use these communities as means for opening the door of inappropriate online activities. Thus, they are providing an open way for cybercrimes such as cyberbullying. In this paper, we deal with the aforementioned issue as a time series modelling methodology, aiming at the recognition of bullying patterns within the questions posed by a predator to his victims. Given a set of real world transcripts (i.e. the whole set of predator*'s* questions), in which each question is numerically labelled in terms of severity, we first model each set of predator*'s* questions as a time series. The next step is the main contribution of this paper, in terms of changing the representation scheme from time series data into symbolic representation. More specifically, inspired by the Multiple Sequence Alignment (*MSA*) method, commonly used in computational biology for identifying conserved regions of similarity among raw molecular data, we represent the set of signals according to a *SAX* (Symbolic Aggregate approXimation) symbolic representation, transforming each signal into a symbol string. The main rationale behind this adoption lies to the fact that the collected cyberbullying data can be converted to string sequences via *SAX* conversion, which in turn can be aligned, thus revealing conserved temporal patterns or slight variations in the attacking strategies of the predators. Experimental results, based on the clustering improvement of the aforementioned data using the extracted patterns instead of the time series data, justify our claims.

1. Introduction

Cyberbullying is a new phenomenon resulting from the advance of new communication technologies, including the Internet, smart phones etc. In all its forms, cyberbullying combines the devastating effects of in-person bullying with several added issues unique to its technological format. A special case of cyberbullying is the online sexual predation. Online sexual predators are defined as adult online users who use Information and Communication Technologies (ICT), in order to exploit vulnerable children or adolescents for sexual or other abusive purposes. In most of the cases, online sexual predators follow common strategies in the ways they attempt to develop relationships with children. These strategies are mainly organized in four sub-categories: personal information, relationship details, activities, and compliments. The exchange of personal information includes details about the physical location of either the victim or the predator, their age, names, the location of the computer within the house. The exchange of relationship information includes discussions on feelings and attitudes and aim at maintaining, building, and dismantling their relationships with each other, as well as with friends and family members. Activities, which constitute a broad category, is primarily defined as the preferred social behaviors commonly shared by both the predator and the victim. This category includes, among others, music preferences, favorite movies and books, practiced sports and hobbies, as well as favorite foods. Finally, during the compliments step, the predator and the victim share praises for each other*'s* appearance, activities, and personal information. Of course, the aim of the predator is to make the victim view him in a positive, appreciative manner. The primary target of an online sexual predator is to earn the trust of the victim in order to isolate and approach him or her. As soon as the victim shows signs of trust towards the sexual

[*] Corresponding author. Tel.: +302273082261.
E-mail addresses: nekpotha@aegean.gr (N. Potha), mmarag@aegean.gr (M. Maragoudakis), lyras@bio.lmu.de (D. Lyras).

http://dx.doi.org/10.1016/j.knosys.2015.12.021
0950-7051/© 2015 Elsevier B.V. All rights reserved.

predator, the offender begins to groom the minor to accept offers of sexual contact [1].

Thus, communication that functions as grooming does not directly lead to sexual contact, but instead, desensitizes the victim to sexual remarks or foul language. A sexual predation could be prone to success when predators deal with minors who are isolated from support networks, be it by low paternal or maternal relationships or by having very few friends. This information is gleaned through online dialogues by asking questions about the minor′*s* social life, by providing sympathy and support in reaction to their situation, and by questioning the strict rules of the parent. The sexual predator seeks to isolate the victim and then to fill the social gaps in the victim′*s* life as a tool to facilitate abuse and gain control of the victim. When the predator has established the victim's trust, commenced grooming and isolated the minor from support networks, the predator attempts to approach the victim by suggesting that they meet for sexual purposes [2].

Undoubtedly, online sexual predation is a social hazard and there are two main factors that have greatly intensified this phenomenon. The former is the anonymity, in combination with the lack of meaningful supervision in the electronic medium. Anyone can be cyberbullied by a stranger or a close acquaintance without ever being able to tell who the culprit is. The fact that a bully can hide behind an electronic veil, disguising his or her true identity, makes bullies feel empowered to say and do more destructive things than they would do face-to-face when they are interacting with their victims in a screen-to-screen manner. The latter is omnipresence, as online sexual predation follows you home. A student being bullied at school may find refuge in other spaces, but a victim of sexual cyberbullying is connected to his or her tormentors whenever he or she is connected to a cell phone or computer which in certain cases for a large number of teenagers includes most of their time [3].

This work presents a novel framework for pattern discovery in real, sexual offender posts. Unlike previous approaches, which consider a predefined window of the predator′*s* questions to the victim, the current approach models the complete set of predator′*s* questions as a time series thus enabling to further emphasize on the dialogue course [4]. Based on the modelling strategy of transforming textual input (i.e. the posts of the sexual predator) as a time-series, we introduce a novel approach for identifying patterns that appear within the predator′*s* signals collection. At first, a transformation of signal data to symbol series using the *SAX* algorithm is applied. Then, we borrow a method from computational biology, known as *MSA* (Multiple Sequence Alignment). This method is suitable for identifying and aligning regions of similarity among the examined sequences, i.e. patterns. The detailed information about each stage of the framework are described in Section 4.1. Results on the clustering performance denoted that, using either Hierarchical Agglomerative Clustering (HAC) or Bayesian Hierarchical Clustering (BHC), we managed to identify which of the signals are close enough, hence belong to the same cluster, applying various distance metrics on the original signals as well as on the transformed signals, based on the frequency of occurrence of the found patterns. Then, we build a cluster tree (dendrogram), employing the three most common cluster similarity measures: single link (*Min*), complete link (*Max*), average link (*Avg*). All dendrograms are evaluated using two well-known unsupervised clustering criteria, namely cophenetic correlation (*CPCC*) and inconsistency coefficient (*Inc. Coef.*).

Although elaborate approaches have been proposed for a variety of tasks such as the detection of patterns in control signals [5] as well as stock market prediction [6], to the best of our knowledge time series pattern detection has never been applied before for sexual cyberbullying data. The main difficulties deriving from the nature of the problem at question are two-fold:

(i) Multiple meaningful patterns may be contained within the data
(ii) Data can be multi-dimensional

To address these issues, in the present approach we go beyond the traditional time series methods (e.g. modeling via Hidden Markov Models (*HMMs*) [7] or *DTW* [8]) and follow a step-wise approach where initially a *SAX* [9] string sequence is extracted by each time series and subsequently, the resulting sequence data are provided as input to *MSA* algorithm that attempts to reveal regions of similarity between the examined sequences and hence identify hidden patterns within the initial time series data. A thorough analysis of the new findings reveals non-trivial behavioral patterns of the predators and by utilizing these patterns as input attributes, we can perform accurate clustering of the dataset which outperforms standards clustering methods.

The rest of this paper is organized as follows. Section 2 presents previous work in cyberbullying detection and briefly reviews related work in time series mining. Section 3 introduces some theoretical background for the analysis techniques used within the current research. Section 4 describes the process of modeling cyberbullying dialogues as time series, performing the low-dimensional discrete representation of time series and implementing *MSA* for pattern detection in *SAX* strings. Section 5 provides the results of the experimental evaluation phase using the dataset described earlier while Section 6 includes the main conclusions drawn from this study and discusses future work directions.

2. Previous related research

In the present work, we propose a novel framework for identifying patterns within cyberbullying data by initially converting them into *SAX* strings which are subsequently processed by a Multiple Sequence Alignment algorithm that aims to retrieve common patterns within the generated sequences. Although, to the best of our knowledge, the incorporation of time series data mining with *MSA* methods has never been proposed before, both approaches are individually widely used in a variety of applications and domains which are extensively discussed in the following subsections.

2.1. Cyberbullying detection and behavior prediction methods

The detection of cyberbullying is a growing problem in the social web and is becoming a major threat to teenagers and adolescents. In web environments, textual content is often unstructured, informal, and even misspelled, making difficult and even impossible the identification of sex offender cyberbullying behavior. As a result, despite the obvious importance of the automatic sexual cyberbullying detection, very few studies have been dedicated on this task.

Some interesting studies on this issue are discussed in this section. A Sexual Predator Identification competition took place for the first time at PAN-2012 [10]. Given a set of chat logs the participants had to identify the predators among all users in the different conversations or the part (the lines) of the conversations which were the most distinctive of the predator′*s* behavior. In conclusion, it was impossible to identify predators using a unique method but it was necessary the use of different approaches. Moreover the most effective method for identifying distinctive lines of the predator behavior in a chat log appeared to be those based on filtering on a dictionary or LM basis. In 2009, Yin et al. [11] proposed a supervised learning approach for detecting harassment. Combining local features, sentiment features, and contextual features training a model for detecting harassing posts in chat rooms and discussion forums. By employing a *SVM* classifier with the linear kernel and