Social Networks Security Policies

Zeineb Dhouioui, Abdullah Ali Alqahtani and Jalel Akaichi

Abstract Social networks present useful tools for communication and information sharing. While these networks have a considerable impact on users daily life, security issues are various such as privacy defects, threats on publishing personal information, spammers and fraudsters. Consequently, motivated by privacy problems in particular the danger of sexual predators, we seek in this work to present a generic model for security policies that must be followed by social networks users based on sexual predators identification. In order to detect those distrustful users, we use text mining techniques to distinguish suspicious conversations using lexical and behavioral features classification. Experiments are conducted comparing between two machine learning algorithms: support vector machines (SVM) and Nave Bayes (NB).

Keywords Privacy protection · Security policies · Social networks · Predators · Text mining · Machine learning · Support Vector Machine

1 Introduction

A social network is a set of entities that may have relationships and is modeled generally as a graph where nodes refers to entities and edges to relations. While facilitating much needed users requirements, social networks enlarge privacy concerns. Among privacy risks, we can mention fraudsters, spammers malicious users especially sexual predators that target children and teenagers. In fact, according to the report of the National Center for Missing and Exploited Children (NCMEC) in 2008, there is 1 in

Z. Dhouioui (✉)
Bestmod, ISG Tunis Université de Tunis, Tunis, Tunisia
e-mail: dhouioui.zeineb@hotmail.fr

A.A. Alqahtani · J. Akaichi
King Khaled University, Guraiger, Abha, Saudi Arabia
e-mail: aalrabaa@kku.edu.sa

J. Akaichi
e-mail: Jalel.akaichi@kku.edu.sa

G. De Pietro et al. (eds.), *Intelligent Interactive Multimedia Systems and Services 2016*, Smart Innovation, Systems and Technologies 55,
DOI 10.1007/978-3-319-39345-2_34

7 teenagers are approached for sexual purposes. The widespread use of the internet and the lack of parental control have brought the cyber-crime and social networks are considered as a new opportunities for pedophiles [3] Social networks are a reach source for sexual predators since these sites are popular for teenagers. In fact, online sexual solicitation raise relatively to the degree of privacy of user-generated content such as profile, photos or statutes [10].

Additionally, Instant messaging is a part of everyday life habits. Thus, the huge flow of instant messages in social networks affect the accuracy of the existing methods proposed to detect sexual predators. Another challenging task is that conversations content is always inaccessible. Moreover, malicious users mask their real identity since it is easy to give false personal information in social networks.

The ability to detect suspicious conversations can help in the improvement of security policies and the detection of such offenders is a primordial security issue. In this work, we present a method for picking out malicious users and identifying suspicious conversations from a set of conversations. Lexical and behavioral features are used to flag sexual predators. Finally, we experiment with SVM and Naive Bayes classification to filter conversations. Ultimately, we find that SVM classifier outperformed the Naive Bayes. The rest of this paper is organized as follows: in Sect. 2 we review existing approaches in this field. Section 3 presents in detail our proposed method based on lexical and behavioral features. The experiments and discussions are outlined in Sect. 4. In Sect. 5, conclusion is depicted and future works are presented.

2 State of the Art

Social networks pedophiles target minor victims [8]. Thus, it is primordial to assure adolescents and promote security policies in social networks mainly by automatically recognizing suspicious conversations. In this section, we focus on briefly reviewing existing methods that handle the identification of sexual predators. Authors in [2] have proposed a two steps approach in order to detect sexual predators using social network dialogues. This approach starts by detecting suspicious conversations where predators participate; then the sexual predators are identified. In [4], sexual predator detection in chat conversation is dealt based on a sequence of classifiers. Indeed, documents are split into three parts according to different stages of predation. To deal with the security issue, Rahman Miah et al. introduce a method able to differentiate child-exploitation, adult-adult and general-chatting dialogues using text categorization approach and psychometric information [9]. In the same way, Bogdanova et al. are convinced that standard text-mining features are relevant to distinguish general-chatting from child-exploitation conversations, but are unable to distinguish between child exploitation and adult-adult conversations [3]. A probabilistic method has been introduced in [7] giving three classes of the chat interventions followed by a sexual predators which are gaining access referring to the intention of predators to approach the victim. Secondly, the deceptive relationship is the preliminary to a sexual attack.

Finally, the sexual affair refers to a sexual affair intention of the predator towards the victim. Since detecting sentiment in texts is helpful to identify online sexual predators, authors in paper [3] present a list of sentiment features. They also used a corpus containing predators chats obtained from http://www.perverted-justice.com. Additionally, authors handle this task via the natural language processing (NLP) techniques. McGhee and his colleagues classified possible sexual predators strategies and introduce the following sexual features [6]:

- Percentage of approach words using verbs like meet, and nouns such hotel
- Percentage of relationship words such as dating words
- Percentage of communicative desensitization words including family members names
- Percentage of words expressing sharing information related to the age, location or also sending photos

The Chat Coder can identify lines in the chat log which include predatory language. The term grooming was defined as a strategy followed by sexual offenders to force their victims to admit the sexual affair. In most cases, the predator tries to isolate the victim in order to easily and more exploit the victim. Actually, grooming is used to describe malicious behavior with the intention of sexual exploitation with a children and is classified as follows: the first class is the communicative desensitization in which vulgar sexual language is used, pornographic con-tents are sent and sexual slang terms are used. The second class refers to the reframing that aims to gain the victim trust by showing online sexual advances. Features extraction [5] is frequently used such as lexical and behavioral ones. Doubtful online text chat can be categorized into two types according to Pendar [8]: interaction between predator and victim and consensual interaction between two adults.

3 The Proposed Method

Defining privacy policies is a challenging concept which requires innovative techniques. Privacy policies are closely related to privacy preserving and users protection and include access control and confidentiality techniques. The new paradigm for security policies require new mechanisms based on trust [1]. Acquiring a high level of trust is primordial before sharing information and build new relationships. Indeed, malicious users appear usually as honest to gain the trust of other users. Users must verify their security parameters and social networks sites are charged with improving security policies to safeguard users from malicious one. Thus, identifying unfaithful users such as sexual predators is a crucial task. Given a social network G(E, V), a member U_i must have a high security level. U_i checks access control mechanisms. Then, he verifies relationships policies based on executing a trust management system based on the detection of sexual predators. Privacy concern includes security, reputation and credibility and finally profiling. In our context, we start by the profiling users to detect credible users to ensure security (Fig. 1).

Fig. 1 The overall framework for security policies

The SPD algorithm (Sexual Predators Detection) can be applied to detect suspect conversation and predators.

Algorithm 1 SPD

Input: conversation sets C
Output: list of sexual predators
1: **while** $C \neq \emptyset$ **do**
2: *featureclassification*(keywords)
3: *Lexicondevelopment*()
4: *Conversationclassification*()
5: *Sum*(*feature*, *weights*)
6: **end while**

To detail the previous algorithm, we describe the following steps:

Step 1: raw data collection: this step consists of collecting conversations
Step 2: Features classification. We distinguish two types of features [3]:

- Behavioral features
  - The number of times a user initiates a conversation
  - The number of questions asked
  - The number of intimate conversations
  - The frequency of turn-taking
  - Intention of grooming or hooking
- Lexical features
  - Percentage of approach words
  - Percentage of relationship words
  - Percentage of communicative desensitization words: these words refer to family members names, for instance, (mm, dad.)
  - Percentage of words expressing sharing information
  - Percentage of isolation
  - Number of emoticons

Step 3: Lexicon development: this step focuses on the informal language of online social networks. For this reason, 3 types of lexicon were created:

- Exchange of personal information
- Grooming
- Approach

Step 4: Suspicious conversation detection: based on features we can classify conversation into two classes positive and negative
Step 5: Flagging users according to predator degree by summing the features weights.

There is no pre-processing stage of conversations texts due to the special characteristics of these conversation texts such as neglecting grammar rules, using abbreviations and emotions. However, pre-filtering is crucial to reduce the computational task by eliminating conversations that contains only one participant or very short ones (less than 10 interventions for both users), or those containing many and several unrecognized characters. Our method can face the challenging task of sexual predators identification as a classification using lexical and behavioral features using a supervised learning. Ideally, we can detect suspicious conversations and distinguish between the victim and the predator. We aimed to anticipate the detection of sexual offenders that can approach teenagers reducing the number of victims and improving social networks security policies. We hypothesize that features are weighted and hence the predaterhood score is the sum of these weights. We believe that weights are appropriate in reflecting the danger of sexual predators. The most interesting from this work is the use of features which flag predatory messages and misbehavior and consequently users can be notified to be aware of suspicious users.

4 Experiments

4.1 Metrics

In what follows, we will focus our experimentation on comparing the performance of the proposed classification method using the following metrics:

$$Precision = \frac{a}{a + b}$$

$$Recall = \frac{a}{a + c}$$

$$F{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where:

- a: the number of conversations appropriately assigned
- b: the number of conversations inaccurately assigned
- c: the number of conversations inaccurately rejected

We compared the performances of different feature sets using the most famous and used learning algorithms: Nave Bayes and SVM classifiers. Actually, SVM is used for binary classification with vastly dimensional space and well perform with text classification. Naive Bayes is simple and relevant for nominal features.

4.2 Results and Discussions

For this study, we were able to exploit data from the web site http://www.perverted-justice.com, but we preferred to gather real conversations from Facebook. Real data gathering was the task requiring the greatest effort. Due to the intimacy of these conversations, we have only collected 30 conversations, where 23 conversations will be used for the training model and 7 for the cross validation. The training model contains 17 conversations of sexual predators and 6 of non-sexual predators classified manually. Instant conversations are characterized by a particular vocabulary. The features extraction refers to find the main characteristics of the text. Features are composed of the most frequent expressions in the collected data expressing misbehavior. We assign weights to behavioral and lexical features (Figs. 2 and 3) and we have also computed the frequency of these features in the following figures (Figs. 4 and 5).

Classification algorithms such as Naive Bayes and Support Vector Machine (SVM) are used. For the evaluation, we used: the precision (P), Recall (R) and the F-measure using the weka tool. Figures 6 and 7 show the results obtained for the previously mentioned classifier. The best performance was obtained using the SVM.

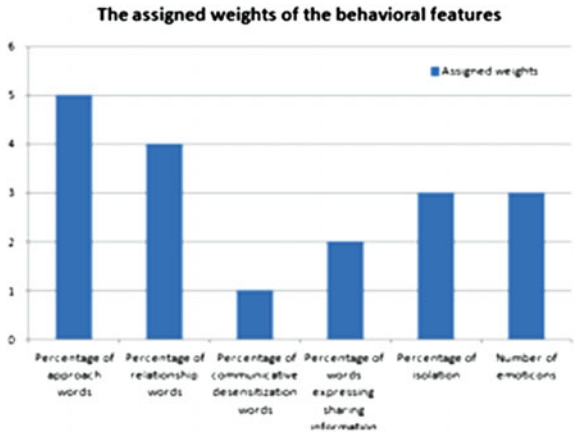

Fig. 2 The assigned weights of the behavioral features

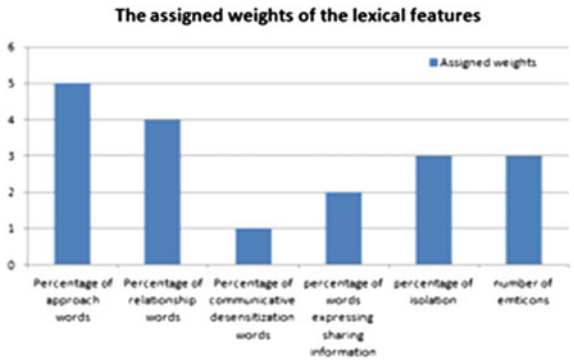

Fig. 3 The assigned weights of the lexical features

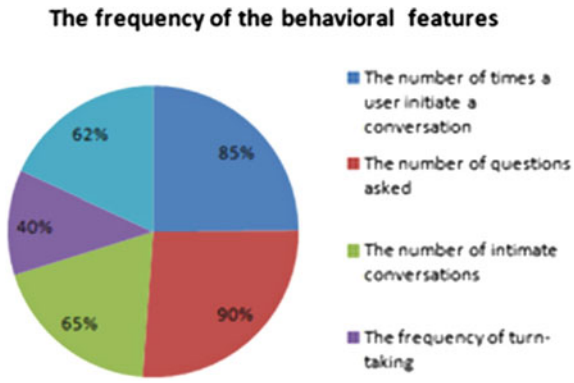

Fig. 4 The frequency of the behavioral features

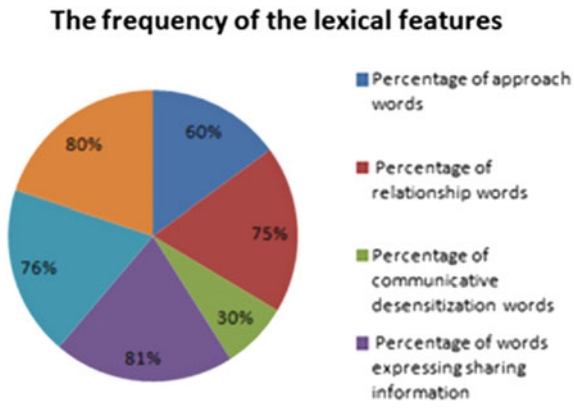

Fig. 5 The frequency of the lexical features

Results can be reported as follows: SVM outperform Naive Bayes for the different settings we considered. Collected data contains both types of conversations, those including sexual predators and conversations between normal users. Therefore, we eliminate conversations with only one participant or containing less than 10 interventions or those including insignificant characters.

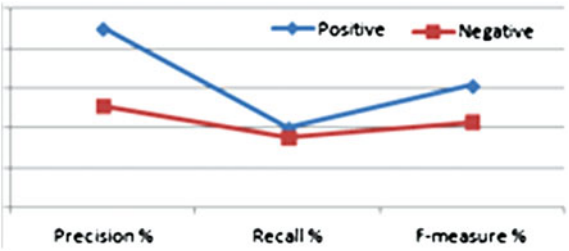

Fig. 6 The accuracy of Naive Bayes classifier

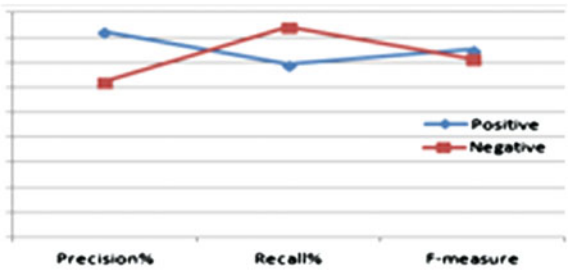

Fig. 7 The accuracy of SVM classifier

Clearly, the best performance was obtained with the SVMs algorithm. Unfortunately, with NB it will be unclear how to detect sexual predators. In fact, conversations has positive classification is 82.2 % with SVM. However, we obtain negative classification is only 62.2 %. Undoubtedly, the F-measure rates shows encouraging results with 75.2 % for Positive and 71.57 % for Negative.

5 Conclusion

One of the main challenges in social networks analysis is to provide high level of privacy protection. While social networks witness a non-stop popularity, real threats raise. For this purpose, this work presents a global overview of some good practice for safe social networks browsing. We discussed possible privacy threats, we have also reviewed existing methods to detect sexual predators and finally based on lexical and behavioral features extracted from texts we can identify suspicious conversations and sexual predators. Experimental results show that SVM classifier present prominent performance than Naive Bayes in terms of precision, recall and the F-measure. Future work includes adding linguistic features and to exploit largest number of conversations.

References

1. Ahn, G.J., Shehab, M., Squicciarini, A.: Security and privacy in social networks. Internet Comput. IEEE **15**(3), 10–12 (2011)
2. Alemán, Y., Vilarino, D., Pinto, D.: Searching sexual predators in social networks
3. Bogdanova, D., Rosso, P., Solorio, T.: Modelling fixated discourse in chats with cyberpedophiles. In: Proceedings of the Workshop on Computational Approaches to Deception Detection, pp. 86–90. Association for Computational Linguistics (2012)
4. Escalante, H.J., LabTL, I., No, L.E.E., ú Villatoro-Tello, E., Juárez, A., Montes-y-Gómez, M.: Sexual predator detection in chats with chained classifiers. In: WASSA 2013, p. 46 (2013)
5. Inches, G., Crestani, F.: Overview of the international sexual predator identification competition at pan-2012. In: CLEF (Online Working Notes/Labs/Workshop), vol. 30 (2012)
6. McGhee, I., Bayzick, J., Kontostathis, A., Edwards, L., McBride, A., Jakubowski, E.: Learning to identify internet sexual predation. Int. J. Electron. Commer. **15**(3), 103–122 (2011)
7. Michalopoulos, D., Mavridis, I.: Utilizing document classification for grooming attack recognition. In: 2011 IEEE Symposium on Computers and Communications (ISCC), pp. 864–869. IEEE (2011)
8. Pendar, N.: Toward spotting the pedophile telling victim from predator in text chats. In: null, pp. 235–241. IEEE (2007)
9. RahmanMiah, M.W., Yearwood, J., Kulkarni, S.: Detection of child exploiting chats from a mixed chat dataset as text classification task. In: Proceedings of the Australian Language Technology Association Workshop, pp. 157–165 (2011)
10. Savirimuthu, J.: Online sexual grooming of children, obscene content and peer victimisation: legal and evidentiary issues. In: Online Child Safety, pp. 61–158. Springer (2012)