Knowledge-Based Systems 000 (2018) 1–14

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts

Ala' M. Al-Zoubi[a,b,*], Hossam Faris[a], Ja'far Alqatawna[a,b,c], Mohammad A. Hassonah[a]

[a] King Abdullah II School for Information Technology, Business Information Technology Department, The University of Jordan, Amman, Jordan
[b] Jordan Information Security and Digital Forensics Research Group (JISDF), Amman, Jordan
[c] Computer Information & Sciences, Higher Colleges of Technology Dubai, UAE

ARTICLE INFO

Article history:
Received 31 December 2017
Revised 25 March 2018
Accepted 21 April 2018
Available online xxx

Keywords:
Spam detection
Feature selection
Social networks
Twitter
Classification
Whale Optimization Algorithm
WOA
Metaheuristic
Support Vector Machine
SVM
Multilingual

ABSTRACT

Detecting spam profiles is considered as one of the most challenging issues in online social networks. The reason is that these profiles are not just a source for unwanted or bad advertisements, but could be a serious threat; as they could initiate malicious activities against other users. Realizing this threat, there is an incremental need for accurate and efficient spam detection models for online social networks. In this paper, a hybrid machine learning model based on Support Vector Machines and one of the recent metaheuristic algorithms called Whale Optimization Algorithm is proposed for the task of identifying spammers in online social networks. The proposed model performs automatic detection of spammers and gives an insight on the most influencing features during the detection process. Moreover, the model is applied and tested on different lingual datasets, where four datasets are collected from Twitter in four languages: Arabic, English, Spanish, and Korean. The experiments and results show that the proposed model outperforms many other algorithms in terms of accuracy, and provides very challenging results in terms of precision, recall, f-measure and AUC. While it also helps in identifying the most influencing features in the detection process.

1. Introduction

Online Social Networks (OSNs) have become very popular as they provide their users with convenient social communication and interactive tools with the ability to instantly publish and share their multimedia contents including video, audio, pictures and rich text. However, these features and the large number of users on these platforms have attracted cybercriminals to exploit such features to perform their malicious activities in an affective and easy way. It was reported that the attacks that used to have a small or limited effect can now have a huge distributed effect through utilizing OSNs sites [1]. OSNs provide attackers the access to a very convenient channel, allowing them to harvest personal information about potential victims and then conducting further social engineering attacks. The attacker can simply analyze the attributes of public profiles and through fake or compromised profile they can send spear phishing messages containing malicious contents to the targeted users. Other aspects of attacks associated with the OSNs are the propagation rate and speed of attacks [2]. For instance, a viral social media message with embedded URL to download a malware might infect thousands of users in a matter of minutes.

While the issue of spam messages is mostly recognized in the context of electronic mailing system, it has become notable on various OSNs where users are heavily suffering from spam. Social spam is the preferred attack vector utilized by cybercriminals to initiate their malicious activities. These unsolicited messages carry deceptive contents with obfuscated URLs to other external sites that might contain malwares, phishing web pages, click-fraud scripts and/or other inappropriate contents. In addition, an early report warned that social media spam is growing faster than comments in most of OSNs sites [3]. The report also shows that 15% of social spam contains links to malicious contents, pornography or malware.

The term Detection was recognized earlier in OSNs by Savage et al. [4] as the profiles that had changed their interactions, or

* Corresponding author at: King Abdullah II School for Information Technology, Business Information Technology Department, The University of Jordan, Amman, Jordan.
E-mail addresses: alaah14@gmail.com (A.M. Al-Zoubi), hossam.faris@ju.edu.jo (H. Faris), j.alqatawna@ju.edu.jo (J. Alqatawna), mohammad.a.hassonah@gmail.com (M.A. Hassonah).

https://doi.org/10.1016/j.knosys.2018.04.025
0950-7051/

might indicate a number of suspected behaviors, such as malicious individuals, which can be spammers, online fraudsters, or sexual predators.

Previous studies indicated that Machine Learning (ML) has played an important role in terms of malicious profile detection in OSNs, using supervised, unsupervised, and semi-supervised techniques. Supervised ML algorithms depend on previously labeled datasets to train themselves and produce classification or prediction models that are then able to predict the label for the new data [5]. On the other hand, unsupervised ML algorithms do not depend on labeling the data, rather, they look for patterns and natural groupings among the instances in the dataset. Semi-supervised algorithms seek smaller portions of labeled data in combination with large portions of unlabeled data; in order to train and build their models.

Realizing the threat and danger of spammers over OSNs, different ML-based models were proposed by researchers in the literature for detecting spammers. For example, [6] proposed a Neural Network (NN) model for detecting spam accounts on Twitter. They used a feedforward type which consisted of activation units [6]. extracted a number of features from the metadata of the tweet and trained their model using Gradient Descent method. The experiments showed that NN outperformed all other approaches with a detection rate of 95.09%. Another work was conducted by Zheng et al. [7], where they developed a model based on Support Vector Machine (SVM) and grid search for detecting spammers on Sina Weibo social network. Feature extraction process was conducted focusing on the content and user-behavior features, then a ranking procedure was applied using Chi-square and InfoGain methods for analysis. The approach showed that it is capable of detecting spammers with an excellent performance against other well-known classifiers. Other works following the same line of research can be found in [8–11]. Most of the previous works focused on developing accurate models for detecting spam and spammers, however, less number of studies focused on identifying the most influencing factors in this identification.

In this work, we propose an ML approach based on SVM classifier and one of the recent metaheuristic optimizers called Whale Optimization Algorithm (WOA). Unlike most of the previous studies, our proposed approach performs automatic detection of spam profiles over OSNs and identification of the most influencing features simultaneously. In this approach, which will be referred to as SVM-WOA, the bubble-net hunting strategy of Whales embedded in the WOA is utilized to optimize the parameters of SVM and search for the best subset of features. The SVM-WOA will be applied on datasets collected from different lingual-contexts from Twitter which are Arabic, English, Spanish, and Korean. The goal is to enhance the detection accuracy and to study the nature of the most influencing features in these lingual contexts. Therefore, the contribution of this work can be summarized in the following three points:

- A new SVM-WOA approach is proposed for accurate detection of spammers over OSNs.
- The SVM-WOA is used to reveal the most influencing features in the process of identifying the spammers.
- Different lingual contexts are studied which are Arabic, English, Spanish, and Korean.

The rest of this paper is structured as follows: Section 2 gives a background of the components used in our proposed technique. While in Section 3, we present the proposed SVM-WOA approach. The data collection and feature extraction processes are detailedly described in Section 4. In Section 5, we list the evaluation criteria for our experiments. Moreover, the findings of this paper are presented and analyzed in Section 6. Finally, we summarize and conclude our work in Section 7.

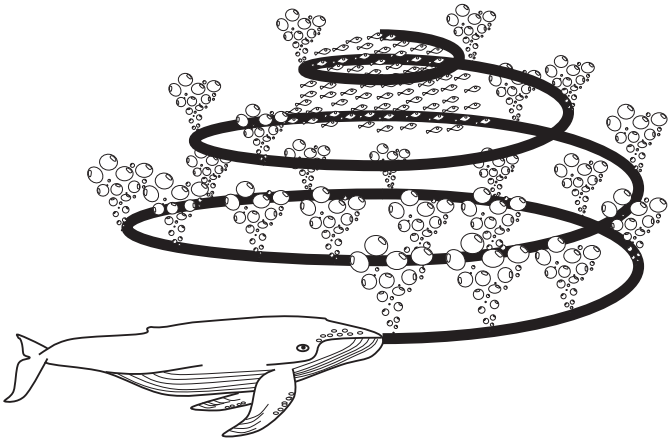

Fig. 1. Bubble-net whale hunting behavior [23].

2. Background

2.1. Feature selection

In Machine Learning, the feature selection technique is a well-recognized process that was applied in much previous work, especially in detecting spam [12–14,49]; in order to reduce the number of features to the most informative and relevant ones; which improves the learning process for the classifier [15–17]. There are two techniques for performing feature selection, filter and wrapper-based techniques. Filter-based technique is a classifier-independent mechanism that uses filter methods to perform a number of statistical measurements to weigh the features and rank them according to their importance [18,19].

In contrast to filters, a wrapper-based method operates on the features in coordination with the classification algorithm to evaluate each subset of features. The quality of the subset can be measured and given by the classifier's selected performance evaluation measure after training the data, and then tested on a hold-out unseen set [20,21,50]. Therefore, wrapper-based feature selection consists of three components: a search algorithm, an inductive algorithm (i.e. a classifier), and an evaluation measure. Although the computation is generally intensive in this method since the classifier should be trained on each subset [22], it leads to better accuracy results than the filter-based methods.

2.2. Whale Optimization Algorithm

Whale Optimization Algorithm (WOA) is a search and optimization algorithm, recently developed by [23]. It is a mathematical simulation of the movement and behavior of humpback whales in their search for food and provisions. The exploitation process of WOA algorithm was inspired by the Bubble-net attacking strategy by whales, where they start targeting fish by forming spiral-shaped bubbles around their fish down to 12m deep from the surface, and then, they swim back up to catch their targeted fish as shown in Fig. 1. The exploration process is represented in this algorithm by the random search of food according to the relative positions of whales; which can be mathematically translated by updating the old solutions through randomly selecting other solutions instead of choosing the best ones. Besides this interesting behavior, WOA is remarkably distinguished from other optimization algorithms by requiring two parameters only to be adjusted. These parameters allow for smooth transition between both the exploration and exploitation processes.

The mathematical representation of WOA algorithm starts with initializing a random set of solutions as shown in Algorithm 1.

Please cite this article as: A.M. Al-Zoubi et al., Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts, Knowledge-Based Systems (2018), https://doi.org/10.1016/j.knosys.2018.04.025