Webology, Volume 17, Number 1, June, 2020

Detecting Fake Accounts on Twitter Social Network Using Multi-Objective Hybrid Feature Selection Approach

Reza Ramzanzadeh Rostami

MSc., Department of Computer Science, Golestan University, Gorgan, Iran. ORCID: 0000-0002-3062-9390. E-mail: ramzanzadeh72@gmail.com

Soheila Karbasi*

*Corresponding author, Assistant Professor, Department of Computer Science, Golestan University, Gorgan, Iran. ORCID: 0000-0001-7334-2229. E-mail: s.karbasi@gu.ac.ir

Received January 26, 2020; Accepted March 20, 2020

Abstract

The frequency of fake accounts or social bots is considered as one of serious challenges of online social networks, which are controlled by automatic operators and often used for malicious purposes. The researchers have performed many efforts to identify these entities in online social networks, which Machine Learning classifier technique using a distinctive feature set is the most common one and feature selection is the principal process of such feature-based classifiers. The present study was carried out the fake accounts detection by using a multi-objective hybrid feature selection approach that helps the feature set selection with optimal classification performance. First, the candidate feature set was identified based on the highest relation to the target class and the least redundancy among the features by the Minimum Redundancy – Maximum Relevance algorithm (mRMR). Then, the stable feature set with the minimum number of features, which can achieve optimal performance, is selected as the final feature set for the detection operations. The proposed approach is tested on two datasets from Twitter's social network and the results were compared to the results of efficient existing methods. According to the results, the performance of the proposed classifier approach is higher compared to existing methods.

Keywords

Online social networks; Fake accounts detection; Feature-based classification; Multi-objective hybrid feature selection; Multi-variate mRMR algorithm

Introduction

Online social networks are popular communication mediums where people share different topics from daily interactions to important events. Various social networks provide many online activities which attract attention of a large number of users, where users increasingly rely on credibility of the information exposed in these networks (Azab, Idrees, Mahmoud, & Hefny, 2016). The amount of attention paid to these networks is increasing with growing popularity of online social networks. Meanwhile, the large amounts of social interactions and user-generated content on these networks make them an attractive environment for researchers of various fields such as sociology, network science, different subcategories of computer science, such as data mining, natural language processing, artificial intelligence, etc. Online social networks are used by organizations in order to promote products and communicate directly with their customers; news media publish their news on these platforms, and these networks are used by politicians and celebrities to connect with their followers (Freitas, Benevenuto, Veloso, & Ghosh, 2016).

The widespread availability and usability are the two most important features of online social networks, which have made them an ideal environment for malicious accounts and cybercriminals. In fact, according to the results of recent studies, online social networks suffer from the publication of a number of automated malicious entities in their environment (Azab et al., 2016). In this study, these automatic malicious entities, which are controlled by computer programs, are called fake accounts (Morstatter, Wu, Nazer, Carley, & Liu, 2016). They simulate the behavior of human users by performing their social activities automatically, such as the sending content and links (Freitas et al., 2016), and used for destructive purposes, such as spreading incorrect information, spamming, spreading rumors and diverting public opinion, promoting malware, fake following, and other malicious purposes in social networks. Confidence in the authenticity of these networks users becomes an essential issue with expanding these automatic accounts in online social networks (Gurajala, White, Hudson, Voter, & Matthews, 2016).

Hypothesizing that most of these accounts are different from social network human users due to the automatic nature of their behaviors (Azab et al., 2016), the proposed detection methods and techniques often use the features of the training set to make a classifier model (Cresci, Di Pietro, Petrocchi, Spognardi, & Tesconi, 2015). Therefore, the feature selection is considered as major process for such feature-based classifiers, whose purpose is to select a minimal set of features that accelerate classification operations and also provide similar or even better performance that

uses all the features for classification operations (Schowe & Morik, 2011). In general, feature selection methods are classified into three groups, including Filter Method, Wrapper Method, and Hybrid Method. The Filter Method evaluates the features based on intrinsic characteristics of the data without utilizing any classification algorithm in such a way that the features are first ranked according to a specific criterion, and then the features with a high ranking are selected for the classification operation. This method is divided into two groups defined as univariate and multivariate. In the univariate technique each feature is ranked independently of its feature space, while the multivariate approach evaluates features in a group, therefore redundancy management is possible (Tang, Alelyani, & Liu, 2014). The filter method can be applied to high dimension data. Some of the most important advantages of this method are high speed and low computational cost. However, this method does not always guarantee the best subset of the feature (Sutha & Tamilselvi, 2015). Wrapper method benefits from the accuracy of a predetermined learning algorithm to identify the quality of the selected features (Tang et al., 2014); this approach compared to the filter method guarantees better results in feature subset selection, but it has a high computational cost and is not preferred for high dimensional data (Sutha & Tamilselvi, 2015). Hybrid methods have been designed to benefit from the advantages of two previous methods. Therefore, like a filter method, the statistical criterion is used for selecting a subset of candidate features aimed to reduce the search space. In the next step, like the Wrapper method, the learning algorithm is used to select the feature subset with high classification performance (Tang et al., 2014).

Despite the fact that feature-based classifier is often used to detect fake accounts in online social networks, but very few attempts have been made to select important features systematically. The approaches used for feature selection process only examine the relationship with the class's feature (Benevenuto, Magno, Rodrigues, & Almeida, 2010; Ahmed & Abulaish, 2013; Azab et al., 2016). While the feature selection based on the highest dependence on the target class, regardless of the relationship between the features can be associated with redundancy and high dependence on the features selected. When two features are highly dependent on each other, if one of them is ignored, the classifier performance for the differentiation will not change (Peng, Long, & Ding, 2005).

Another important issue in feature selection is identifying the number of desirable features for classification operations. Useful information may be lost by deleting a large number of features and in contrast to keeping a large number of features can prevent achieving the desired performance. The features can be selected correctly by checking the performance and stability of the feature set (Kuncheva, 2007; Saeys, Abeel, & Van De Peer, 2008; Meinshausen & Bühlmann, 2010).

The stability of the feature selection method can be calculated by the similarity of the subset of the features created by a feature selection method on different data of the same size. If a small

change in the input data causes a significant change in the selected feature set, then the desired feature set cannot be considered as the final solution (Schowe, 2010). In addition, the computational cost of features is another priority that should be considered in the field of online social networks that deal with a large amount of data (Cresci et al., 2015).

A classification-based approach with an emphasis on the selection of effective and lightweight features is used in the present study to achieve a balanced and desirable performance. In this regard, firstly, the statistical and lightweight features are extracted by reviewing the academic articles presented in this field and the computational cost is considered. Then, a multi-objective hybrid approach is used to select the important features in order to detect fake accounts, while the multivariate mRMR algorithm (Ding & Peng, 2003) is applied to select the candidate feature set. This method considers redundancy among features in addition to examining the relationship of features compared to the target class. In addition, the stability of each candidate feature set is calculated by examining them on different parts of the datasets and three classification algorithm of Random Forest, Naïve Bayes and Support Vector Machine are used to calculate the relevant performance. Finally, the stable feature set with optimal performance is selected as the final feature set for fake accounts detection operations. It is characterized by the minimum number of features. The proposed approach performance was evaluated using two Twitter social networking datasets, and according to the results, the proposed approach could achieve the desired performance for two datasets, when compared to other methods under study.

Related Works

Online social networks are the ideal environment for creating fake accounts due to growing popularity of these networks. Numerous studies have been carried out to identify these automatic entities in online social networks. Romanov et al. proposed a review of the state-of-the-art publications dedicated to detection of fake profiles in social media (Romanov, Semenov, Mazhelis, & Veijalainen, 2017). The article explains the important role of fake identities in advanced persistent threats and covers two main approaches of detecting fake social media accounts; the approaches which aim to analyze individual accounts and the approaches which capture the corresponding activities spanning a large group of accounts. Most of the presented approaches use the machine learning technique to detect fake accounts. Chu et al. analyzed behavioral characteristics of social bots and human users on the Twitter's social network and a classifier was used to classify the accounts (Chu, Gianvecchio, Wang, & Jajodia, 2012). Link prediction in social networks has potential applications such as discovering faked connections. As well as social networks often have faked links, their discrimination will help to improve their performance (Hasan & Zaki, 2011).

Jalili et al. (2017) studied the link prediction problem in complex Twitter networks and two layers of the connectivity information are used to predict the links. These experiments show that,

the networks are not highly correlated in the layers; nevertheless, the cross-layer information significantly improves the prediction performance. Three classical classifiers were used for the classification tasks with SVM classifier displaying the best performance of (Jalili, Orouskhani, Asgari, Alipourfard, & Perc, 2017).

Davis et al. introduced a random forest based classifier with over 1000 features in order to detect fake accounts (Davis, Varol, Ferrara, Flammini, & Menczer, 2016). Although, Yang et al. designed a number of robust features by performing empirical analysis of the evasion tactics utilized by fake accounts (Yang, Harkreader, & Gu, 2013). Fazil and Abulaish introduced four categories of different features for classification of social bots and human users on the Twitter social network, based on the assumption that fake accounts do not control the quality of their followers and often follow them in a blindly manner (Fazil & Abulaish, 2018). Cresci et al. analyzed features discussed in academic papers and online websites as indicative for fake accounts detection on the Twitter social network, and features with the best performance in terms of cost and detection efficiency in classification operations (Cresci et al., 2015). They created and kept automatic fake accounts and found that identification of the profile templates was the fast method for discrimination of these automated entities without analyzing the content of tweets. Gurajala et al. (2016) applied the pattern analysis of the usernames and distribution of update times in the Twitter social network accounts for early detection of these automated entities.

Results of recent studies show that fake accounts evolved over time and use advanced techniques in their social communication and behavioral patterns (Cresci, Spognardi, Petrocchi, Tesconi, & Pietro, 2019). Therefore, analysis of profile information, regardless the content sent does not guarantee accurate and efficient detection.

Atodiresei et al. presented a method of fake users and fake news identification in Twitter social network, which approved that detecting whether the news is a fake, or not, might be inefficient based alone on its popularity in a social network (Atodiresei, Tănăselea, & Iftene, 2018). This method receives the link to a tweet and computes its credibility, based on comparison to trustworthy news sources and on the credibility of the user. It also computes other statistics, such as the overall emotion (sentiment) of the tweet, taking into consideration the emojis and hashtags used in the Twitter.

Mohammadrezaei et al. (2018) presented a new approach, which was based on similarity between the networks of users' friends in order to discover fake accounts in Twitter social networks (Mohammadrezaei, Shiri, & Rahmani, 2018). Similarity measures such as common friends, cosine, jaccard and l1-measure were calculated from the adjacency matrix of the corresponding graph of Twitter social network. The experiments show that the Medium Gaussian SVM algorithm predicts fake accounts with higher accuracy. While most of methodologies use

classification method for fake accounts detection, Cresci et al. and Miller et al. proposed different approaches (Miller, Dickinson, Deitrick, Hu, & Wang, 2014; Cresci, Di Pietro, Petrocchi, Spognardi, & Tesconi, 2016). In particular, the online activities modeling method introduced by Cresci et al., is inspired by the biological DNA concept to detect automated entities, in such a way that online users' activities are encoded in form of the sequence of character strings and similar sequences are used for social bots detection (Cresci et al., 2016). Miller et al. (2014) introduced the Anomaly Detection Method for fake accounts detection on Twitter, which makes a clustering model for human users, then all the outlier points are considered as fake accounts.

Despite the widespread use of classification technique for fake account detection in online social networks, the process of effective feature selection is rarely considered. The GAIN univariate algorithm was used by El Azab et al. to evaluate the importance of the features in fake accounts detection, which determines the weight of each feature based on its importance (Azab et al., 2016). Benevenuto et al. used the Chi-Square criterion to identify effective features (Benevenuto et al., 2010). Ahmed and Abulaish (2013) presented that elimination of high-importance features reduces the classifier performance significantly.

Ojo proposed a method, which can accurately identify fake profiles in online social networks by application of the Natural Language Processing technique to eliminate or reduce the size of the dataset resulted in significant improvement of the overall model performance (Ojo, 2019). Principal Component Analysis (PCA) is used for appropriate feature selection. After extraction, six features were found which influenced the classifier. Support Vector Machine (SVM), Naïve Bayes and Improved Support Vector Machine (ISVM) were used as classifiers. ISVM introduced a penalty parameter to the standard SVM objective function to reduce the inequality constraints between the slack variables, which conducted a better result.

It is obvious that, in reviewed approaches, the feature selection process is performed only by consideration the feature relationships in comparison to the target class, while dependency between the features is ignored. In some later studies, researches considered stability of the feature set which is also important issue. However, stability alone is not a sufficient measure for evaluation of the feature set, accordingly, it was often considered in a combination with classifier performance (Saeys et al., 2008). Therefore, the issue of the number of desirable features, which can be selected for the detection operation by examining the performance and stability of the feature set, is considered in the proposed approach, as explained below.

Proposed methodology

This study is carried out to detect fake accounts on Twitter social network by using an effective and lightweight feature set that provides optimal performance and accuracy. The used approach

has two steps of extracting statistical features and selection important features using the multi-objective hybrid method. Therefore, in the first subsection, the extracted features are described.

1. Detection features

User profile, contents and social connections, are considered as three useful information sources for creating features in online social networks (Davis et al., 2016). In the present study, profile information and tweet contents on the Twitter social network is used for extraction of 46 statistical features. At first glance, the number of features may raise concerns on the computational cost, but it should be noted that many of these features have been obtained from evaluation of standard statistical criteria of other features, such as entropy and standard deviation. The extracted features have been identified by examining academic articles in this field and considering computational cost. Actually, we have extracted features that represent different aspects of behavior of Twitter accounts, and there is no need in heavy processing for extraction, which increases costs of operations. Table 1, shows these 46 features.

Table 1. Extracted statistical features from profile information and tweets content

Features Description	
Follower count	Hashtag count
Follower count/Account Age [1]	Hashtag count/Tweet count
Following count	Hashtag-per-tweet Standard deviation
Following count/Account Age	Hashtag-per-tweet Entropy
Follower count/Following count	Tweets-with-Hashtags proportion
Follower count/Following count^2	The consecutive tweets interval mean
(2×Follower count)−Following count	The consecutive tweets interval Standard deviation
Follower count/Follower + Following	Link count
Favorites count	Link count/Tweet count
Favorites count/Account Age	Link-per-tweet standard deviation
Tweet count	Link-per-tweet Entropy
Tweet count/Account Age	Tweets-with-Links proportion
List count	Mention count
List count/Account Age	Mention count/Tweet count
Favorites count/Tweet count	Mention-per-tweet Standard deviation
List count/Follower count	Mention-per-tweet Entropy
GEO Tag	Tweets-with-Mentions proportion
Retweet count	Reply count
Retweet count/Tweet count	The consecutive Replies interval mean
The consecutive Retweets interval mean	The consecutive Replies interval standard deviation
The consecutive Retweets interval Standard deviation	Reply count/Mention count
The number of times the tweets sent by the user are retweeted by other users	The total number of the received likes
	Received likes count/Tweet count
	Received like-per-tweet Standard deviation

1 Account age= the number of days from the date of creation the account and the date of collecting the dataset.

2. Important features selection

In this subsection, the proposed approach to select the effective features for fake accounts detection is described. As mentioned earlier, in the proposed approach, a compact set of superior features is selected to reduce search space by mRMR algorithm (Ding & Peng, 2003). Assume that F indicates a feature set available, which includes $P = \|F\|$ features; the mRMR algorithm selects a feature with the highest correlation to the target class (y), which is shown in Formula 1.

$$F_1 = \arg\max_{x_i} \left(cor(x_i, y)\right) \quad x_i \in F \quad i=1,\ldots,p \tag{1}$$

Assume that F_j indicates the selected feature set in the step $j \in \{1, .., K\}$ of the mRMR algorithm. According to Formula 2, the algorithm adds a feature to F_j using a repetitive process, which comprises the best correlation and redundancy rates compared to other features.

$$F_{j+1} = F_j \cup \left\{ \arg\max_{f \in F/F_j} \frac{Cor(f,y)}{\frac{1}{j}\sum_{g \in F_j} Cor(f,g)} \right\} \tag{2}$$

In order to calculate the correlation of the features to the target class, two criteria are used by the mRMR algorithm in accordance with the type of the feature under study. The mutual information, which is a common criterion for defining the dependence of variables and used for discrete / nominal features (Formula 3) and the F-test measure that evaluates the rate of variance between the class and the mean of the variance within the class that is provided for continuous variables (Formula 4).

$$I(x,y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{3}$$

$$F(x,y) = \frac{(n-C)\sum_c n_c (\bar{x}_c - \bar{x})^2}{(C-1)\sum_c (n_c - 1)\sigma_c^2} \tag{4}$$

As mentioned earlier, the mRMR algorithm considers the dependency between the features in addition to evaluation of the relationship to the target class. Similarly, the mutual information criterion is used to calculate the redundancy of the discrete / nominal features, while the redundancy of continuous features is calculated using the Pearson's correlation coefficient (Formula 5).

$$R(x_i, x_j) = \frac{Cov(x_i, x_j)}{\sqrt{Var(x_i) Var(x_j)}} \quad x_i, x_j \in F \tag{5}$$

In order to identify the candidate feature set, stratified sampling is used to divide the dataset into K subset with the same size and the mRMR algorithm is applied to all of these subsets. In addition, the mean weights obtained by the mRMR algorithm for K subset of the dataset are used

to calculate the weight of each feature. At the end, p candidate feature sets are identified using the mRMR algorithm, so that $F_i(\forall i \in [1,p])$ indicates the candidate feature set including i superior feature identified by this algorithm.

The stability of each candidate feature set is calculated using a similarity based approach. Based on this approach, the stability of the feature selection process is calculated using the mean similarity of all pairs of feature sets created on K subsampling of the dataset, as it is shown by Formula 6. F_i feature sets selected on the subsampling i and $S(F_i,F_j)$ shows the similarity criterion of the pairs of features F_i and F_j. The Jaccard Index criterion is used to calculate the similarity of the pairs of feature sets created on K subsampling of the dataset (Formula 7) (Saeys et al., 2008).

$$S_{tot}=\frac{2\times\sum_{i=1}^{k}\sum_{j=i+1}^{k}S(F_i,F_j)}{k\times(k-1)} \quad (6)$$

$$S(F_i,F_j)=\frac{|F_i\cap F_j|}{|F_i\cup F_j|} \quad (7)$$

The performance of each candidate feature set is evaluated using three classifiers algorithms: Random Forest, Naïve Bayes and Support Vector Machine. In this regard, the criteria of classification accuracy, F-Measure and Matthews correlation coefficient (MCC) have been used. According to accuracy or classification rate, it can be concluded that how many percent of the total experimental dataset is properly categorized by classification algorithm (Giudici & Figini, 2009) which this concept is shown by Formula 8.

$$\text{Accuracy}=\frac{TP+TN}{TN+FN+TP+FP} \quad (8)$$

However, classification accuracy alone is not considered as a good measure for classifier performance evaluation in identifying fake accounts, because the same values are considered for the accounts of different classes (Giudici & Figini, 2009); it is obvious that, fake accounts detection is more important than human users detection. In addition, the precision and recall are considered as important criteria (Morstatter et al., 2016). Therefore, along with the classification accuracy, F-Measure is also used to evaluate the performance of the candidate feature set. This criterion is used when it is not possible to assign the importance to any of two criteria of precision and recall. Formula 9 shows the definition of F-Measure.

$$\text{F-Measure}=\frac{2\times\text{Recall}\times\text{Precision}}{\text{Recall}+\text{Precision}} \quad (9)$$

The MCC is also one of the best assessment criteria, which explains the confusion matrix in the form of numbers. This criterion indicates the correlation coefficient between the predicted classification by the model compared to the actual class, which is given in Formula 10 (Nellore, 2015).

$$MCC=\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (10)$$

Finally, the proposed approach selects the last feature set which has optimal classification performance and includes a minimum number of features and shows high stability in the subsampling of the dataset.

Results

Two datasets of Twitter social network[2] (Cresci et al., 2019) are used in this section and their description is presented in Table 2. The Social Spam Bot 1 includes fake accounts which were used by one of the municipality candidates in Rome to publish its policies in 2014; the Social Spam Bot 3 also includes bots that are used to advertise and publish the Amazon company products. The Genuine accounts also refer to human users in the datasets.

The RapidMiner software is used to implement the tests. In addition, *K* subsampling with the size of 40 are extracted from the datasets to identify and calculate the stability of the candidate feature set. Meanwhile, 10-Fold cross validation is used to calculate the performance criteria of the classifier algorithms.

Table 2. Characteristics of used datasets

Dataset	Description	Accounts	Tweets
Test Set 1	Genuine accounts + Social Spam Bot 1	1982	4061598
Test Set 2	Genuine accounts + Social Spam Bot 3	928	2628181

1. Evaluation results

Table 3, shows the results of applying all 46 extracted features on the Test Set 1 and Test Set 2 using three algorithms of Random Forest, Naïve Bayes and Support Vector Machine.

Table 3. Classification performance using all 46 extracted features on Test Set 1 and Test Set 2

Dataset	Machine learning algorithm	Accuracy%	F-Measure%	MCC%
Test Set 1	Random Forest	97.7	97.7	95.47
	Naïve Bayes	96.6	96.4	93.31
	SVM	97.9	97.9	95.76
Test Set 2	Random Forest	97	96.9	94.01
	Naïve Bayes	97.1	97	94.26
	SVM	96.5	96.6	93.13

2 Available at: http://mib.projects.iit.cnr.it/dataset.html

The proposed approach identifies the stable feature set (candidate feature set identified by mRMR algorithm) with an optimal classification performance. It also has the minimum feature number for each of three classifier algorithms (the results are listed in Table 4). According to the results, the proposed approach for the Test Set 1 and Test Set 2 datasets could achieve better performance with a minimal feature set when compared to the analyzes where all 46 features was considered for detection operations.

Table 4. Classification performance using the feature set identified by proposed approach

Dataset	Machine learning algorithm	No. features	Robustness %	Accuracy %	F-Measure %	MCC %
Test Set 1	Random Forest	8	100	98	98	95.98
	Naïve Bayes	8	100	97.6	97.5	95.21
	SVM	8	100	98	98	96.06
Test Set 2	Random Forest	17	100	97.1	97	94.21
	Naïve Bayes	15	100	97.2	97.1	94.54
	SVM	7	100	97.1	97.1	94.19

In the Test Set 1, the algorithms of Random Forest and SVM could achieve optimal results which are close to each other. As mentioned earlier, achieving balanced performance is considered here as one of the major objectives of the proposed approach. According to the results of reviewing performance criteria in more detail (Table 5), the Support Vector Machine algorithm could achieve balanced solutions if compared to Random Forest. In particular, the result of 98% has achieved for most of performance criteria. In the Test Set 2, the Naïve Bayes algorithm achieved the best performance when fifteen features were applied. The Support Vector Machine algorithm has achieved very close results compared to the Naïve Bayes algorithm using seven features. Therefore, the Support Vector Machine algorithm has been selected as the best classifier for this dataset.

Table 5. Performance of Random Forest and SVM algorithm on Test Set 1

Machine learning algorithm	Precision %	Recall %	Specificity %	Accuracy %	F-Measure %	MCC %
Random Forest (8 Features)	98.8	97.2	98.8	98	98	95.98
SVM (8 Features)	98	98.1	98	98	98	96.06

The feature set with the best performance processed with the Support Vector Machine algorithm for two datasets is listed in Table 6 and the obtained results were compared to the methods discussed by (Cresci et al., 2019) as it is shown in Table 7. The BotOrNot system (Davis et al., 2016), while it evaluates more than a thousand Twitter account features for identification, shows

inefficient performance in fake account detection from both datasets. Its performance for Test Set 1 (F-Measure = 28.8% and MCC = 17.4% respectively) shows that BotOrNot tends to classify the fake accounts of this dataset as genuine accounts.

As it is shown in Table 7, the method proposed by Yang et al. (Yang et al., 2013) failed to detect the bots in the datasets due to its low recall. 126 profile-based features and tweets content are used when Miller et al. (Miller et al., 2014) method was applied to detect the fake accounts from the two datasets with insufficient outcome. Low values for precision and recall show that detection operations by this method are unreliable.

The proposed approach without heavy processing of the content of tweets has achieved optimal and balanced performance using only eight features for Test Set 1 and seven features for Test Set 2.

Table 6. The feature set selected by proposed approach for Test Set 1 and Test Set 2

Dataset	Selected Features
Test Set 1	GEO Tag
	Retweet count/Tweet count
	Mention count
	Mention count/Tweet count
	Mention-per-tweet standard deviation
	Mention-per-tweet Entropy
	Tweets-with-Mentions proportion
	Reply count/Mention count
Test Set 2	Retweet count/Tweet count
	Link count
	Mention count/Tweet count
	Mention-per-tweet standard deviation
	Mention-per-tweet Entropy
	Tweets-with-Mentions proportion
	Reply count/Mention count

Table 7. Investigating the results of proposed approach compared to the methods discussed in (Cresci et al., 2019)

Dataset	Approaches	Precision %	Recall %	Specificity %	Accuracy %	F-Measure %	MCC %
Test Set 1	Proposed approach	98	98.1	98	98	98	96.06
	(Davis et al., 2016)	47.1	20.8	91.8	73.4	28.8	17.4
	(Yang et al., 2013)	56.3	17	86	50.6	26.1	4.3
	(Miller et al., 2014)	55.5	35.8	69.8	52.6	43.5	5.9
	(Ahmed & Abulaish, 2013)*	94.5	94.4	94.5	94.3	94.4	88.6
	(Cresci et al., 2016)	98.2	97.2	98.1	97.6	97.7	95.2
Test Set 2	Proposed approach	96.5	97.9	96.4	97.1	97.1	94.19
	(Davis et al., 2016)	63.5	95	98.1	92.2	76.1	73.8
	(Yang et al., 2013)	72.7	40.9	84.8	62.9	52.4	28.7
	(Miller et al., 2014)	46.7	30.6	65.4	48.1	37	−4.3
	(Ahmed & Abulaish, 2013)*	91.3	93.5	91.2	92.3	92.3	84.7
	(Cresci et al., 2016)	100	85.8	100	92.9	92.3	86.7

* Modified by (24): employing Fast greedy instead of MCL for graph clustering step

The Graph clustering and Community Detection methods were used by Ahmed and Abulaish to detect fake accounts (Ahmed & Abulaish, 2013). However, according to the results of the studies conducted by Cresci et al. (Cresci et al., 2019), this approach is associated with errors. When these methods were applied to our data sets, all Test Set 1 and Test Set 2 accounts were assigned to a single cluster. The Fast Greedy Community Detection algorithm is replaced with the Markov clustering algorithm (MCL) to solve this problem. The results obtained for this adjusted version are listed in Table 7, which shows better performance for the same two datasets when compared with the alternative method.

The approach proposed by Cresci et al. (Cresci et al., 2016), could not achieve balanced and desirable results. The results of precision and recall obtained from this approach are unbalanced, which reduces the accuracy of the classifier (F-Measure and MCC). Our proposed method, considers the balance between performance criteria, and achieves results in a desirable classification performance compared to other methods are mentioned above.

2. Analysis of selected features by the proposed approach

In this subsection, the selected feature set is analyzed and examined using Test Set 1 and Test Set 2. As it is shown on Table 6, the Mention-related features are important for two datasets. Figure 1 and Figure 2 show the Cumulative Distribution Function (CDF) of the Mention-related features that are selected by the proposed approach for two datasets.

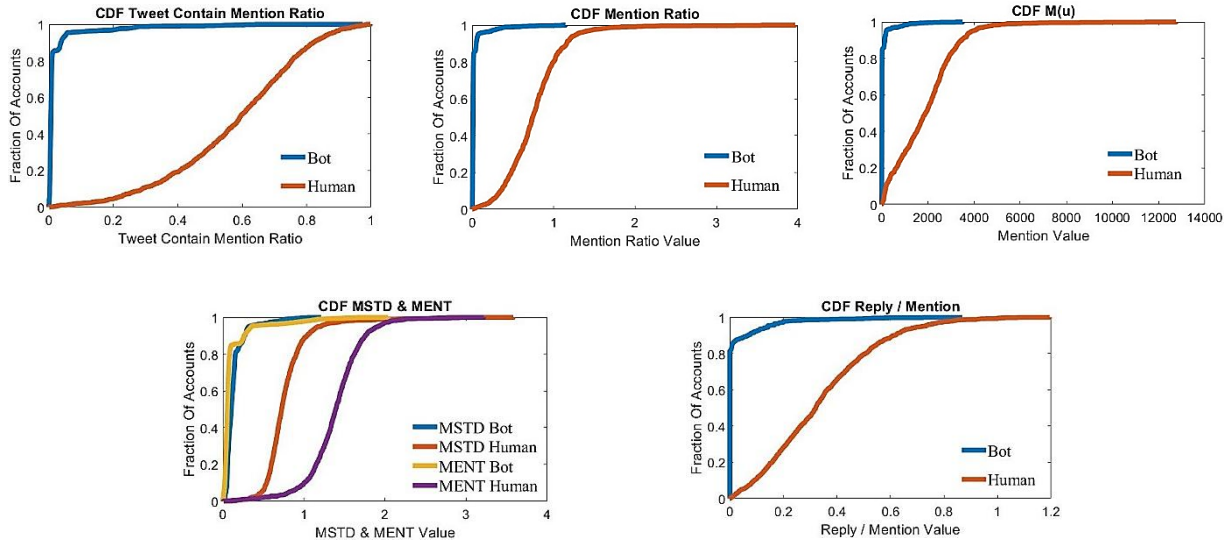

Figure 1. Cumulative Distribution Function of Mention related features for Test Set 1

As it is shown in Figure 1 and Figure 2, fake accounts of two datasets show similar behaviors to each other, but their behaviors are different from human users, who use Mention for their tweets. These automatic entities follow the same pattern in the use of Mention in their tweets, while a specific pattern is not followed by human users and they are significantly more diverse when compared with fake accounts. As it was demonstrated by Fazil and Abulaish, the fake accounts often use Mention in their tweets to increase the similarities to human users, and thus show high rates for this feature (Fazil & Abulaish, 2018); nevertheless, the fake accounts from two datasets have lower Mention numbers and Mention rates than human users.

The retweet rate of accounts is considered as another feature selected by the proposed approach for two datasets. Fake accounts are not smart enough to simulate the content generation behavior of human users. These automated entities often retweet the tweets sent by other users to send content, or create tweets using probabilistic methods, such as the Markov chain algorithm, or use specific databases for sending content. Therefore, it is expected that these automatic entities provide a higher rate of retweet compared to human users (Fazil & Abulaish, 2018). However, the fake accounts of the datasets under study had completely different behaviors, as shown in Figure 3, which means existing bots rarely used retweet to send content.

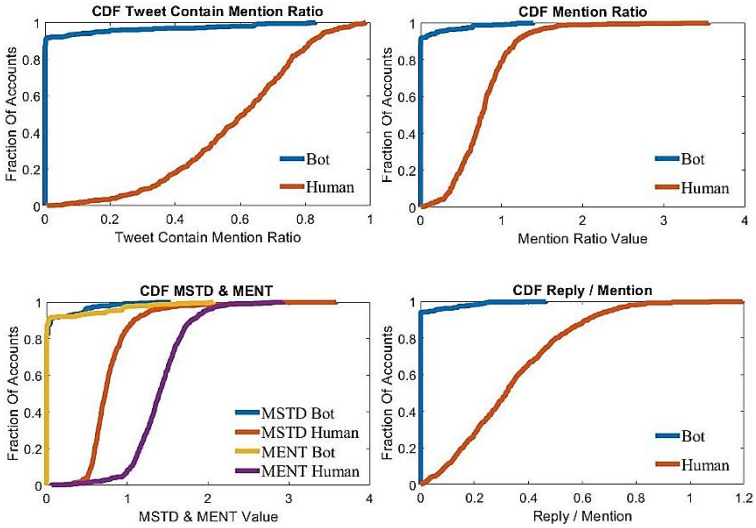

Figure 2. Cumulative Distribution Function of Mention related features for Test Set 2

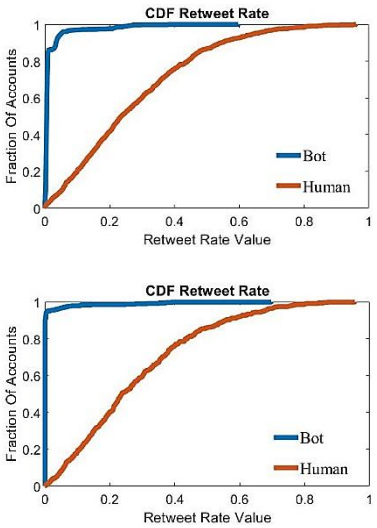

Figure 3. Cumulative Distribution Function of Retweet Rate feature for two datasets

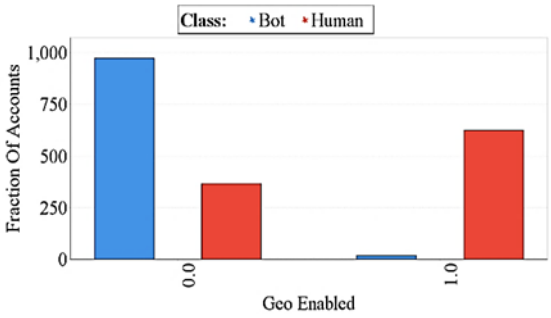

Figure 4. Histogram of GEO Tag feature for Test Set 1

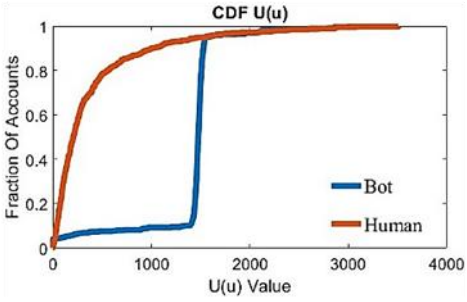

Figure 5. Cumulative Distribution Function of link count for Test Set 2

The *GEO Tag feature* for the Test Set 1 is shown in Figure 4. As shown in this figure, fake accounts of this dataset often disabled *GEO Tag feature* in their tweets (98% ~), while 63.7% of human users enabled this feature. Figure 5 shows the Cumulative Distribution Graph for the number of links used in the tweets for Test Set 2 dataset. The number of links used by the fake accounts of this dataset is higher than of human users, which is expected, because the robot accounts of this dataset are responsible for advertising the products of Amazon Company, which is resulted in the excessive use of links in the tweets of these automated entities compared to human accounts.

Conclusion

The feature selection process in classification-based approaches is very important to detect fake accounts in online social networks, because the classifier system should be applied on a high volume of data and in a real time. In addition, identified features are often analyzed and investigated by researchers in order to understand behavior of fake accounts. It means that

stability of the feature selection process is another priority, which should be considered. Also, stability alone is not considered as a sufficient measure for the selected features evaluation and is often combined with classification performance. In this study, a multi-objective hybrid approach is used to identify the effective feature set for fake accounts detection in the Twitter social network. The outcomes of experiments performed on two datasets from Twitter demonstrated that the proposed approach could achieve more optimal and balanced performance when compared to other existing methods. The proposed approach can be applied for various social networks with a small change in the feature set for detection of fake accounts, which can be an objective for future studies.

References

Ahmed, F., & Abulaish, M. (2013). A generic statistical approach for spam detection in Online Social Networks. *Computer Communications*, *36*(10–11), 1120–1129. DOI: 10.1016/j.comcom.2013.04.004

Atodiresei, C. S., Tănăselea, A., & Iftene, A. (2018). Identifying Fake News and Fake Users on Twitter. *Procedia Computer Science*, *126*, 451–461. https://doi.org/10.1016/j.procS.2018.07.279

Azab, A. El, Idrees, A. M., Mahmoud, M. A., & Hefny, H. (2016). Fake account detection in twitter based on minimum weighted feature set. International Journal of Computer, Electrical, Automation, Control and Information Engineering. Retrieved January 10, 2020, from https://www.researchgate.net/publication/304569053_Fake_Account_Detection_in_Twitter_Based_on_Minimum_Weighted_Feature_set

Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting spammers on Twitter. In *7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference, CEAS 2010*.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*, *9*(6), 811–824. https://doi.org/10.1109/TDSC.2012.75

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems*, *80*, 56–71. https://doi.org/10.1016/j.dss.2015.09.003

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2016). DNA-Inspired Online Behavioral Modeling and Its Application to Spambot Detection. In *IEEE Intelligent Systems* (Vol. 31). https://doi.org/10.1109/MIS.2016.29

Cresci, S., Spognardi, A., Petrocchi, M., Tesconi, M., & Pietro, R. Di. (2019). The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. *26th International World Wide Web Conference 2017, WWW 2017 Companion*, 963–972. DOI: 10.1145/3041021.3055135

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). *BotOrNot: A System to Evaluate Social Bots*. 4–5. https://doi.org/10.1145/2872518.2889302

Ding, C., & Peng, H. (2003). Minimum redundancy feature selection from microarray gene expression data. *Proceedings of the 2003 IEEE Bioinformatics Conference, CSB 2003*, 523–528. https://doi.org/10.1109/CSB.2003.1227396

Fazil, M., & Abulaish, M. (2018). A Hybrid Approach for Detecting Automated Spammers in Twitter. *IEEE Transactions on Information Forensics and Security*, *13*(11), 2707–2719. https://doi.org/10.1109/TIFS.2018.2825958

Freitas, C., Benevenuto, F., Veloso, A., & Ghosh, S. (2016). An empirical study of socialbot infiltration strategies in the twitter social network. *Social Network Analysis and Mining*, *6*(1). https://doi.org/10.1007/s13278-016-0331-3

Giudici, P., & Figini, S. (2009). Applied Data Mining for Business and Industry. In *Applied Data Mining for Business and Industry*. https://doi.org/10.1002/9780470745830

Gurajala, S., White, J. S., Hudson, B., Voter, B. R., & Matthews, J. N. (2016). Profile characteristics of fake Twitter accounts. *Big Data and Society*, *3*(2), 205395171667423. https://doi.org/10.1177/2053951716674236

Hasan, M. Al, & Zaki, M. J. (2011). A Survey of Link Prediction in Social Networks. In *Social Network Data Analytics* (pp. 243–275). https://doi.org/10.1007/978-1-4419-8462-3_9

Jalili, M., Orouskhani, Y., Asgari, M., Alipourfard, N., & Perc, M. (2017). Link prediction in multiplex online social networks. *Royal Society Open Science*, *4*(2). https://doi.org/10.1098/rsos.160863

K. Ojo, A. (2019). Improved Model for Detecting Fake Profiles in Online Social Network: A Case Study of Twitter. *Journal of Advances in Mathematics and Computer Science*, 1–17. https://doi.org/10.9734/jamcs/2019/v33i430187

Kuncheva, L. I. (2007). A stability index for feature selection. In *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, AIA 2007*.

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *72*(4), 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

Miller, Z., Dickinson, B., Deitrick, W., Hu, W., & Wang, A. H. (2014). Twitter spammer detection using data stream clustering. *Information Sciences*, *260*, 64–73. https://doi.org/10.1016/j.ins.2013.11.016

Mohammadrezaei, M., Shiri, M. E., & Rahmani, A. M. (2018). Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms. *Security and Communication Networks*, *2018*. https://doi.org/10.1155/2018/5923156

Morstatter, F., Wu, L., Nazer, T. H., Carley, K. M., & Liu, H. (2016). A new approach to bot detection: Striking the balance between precision and recall. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 533–540. https://doi.org/10.1109/ASONAM.2016.7752287

Nellore, S. B. (2015). Various performance measures in Binary classification-An Overview of ROC study. In *IJISET-International Journal of Innovative Science, Engineering & Technology* (Vol. 2). Retrieved January 10, 2020, from www.ijiset.com

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238. https://doi.org/10.1109/TPAMI.2005.159

Romanov, A., Semenov, A., Mazhelis, O., & Veijalainen, J. (2017). Detection of fake profiles in social media: Literature review. *WEBIST 2017 - Proceedings of the 13th International Conference on Web Information Systems and Technologies*, 363–369. https://doi.org/10.5220/0006362103630369

Saeys, Y., Abeel, T., & Van De Peer, Y. (2008). Robust feature selection using ensemble feature selection techniques. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5212 LNAI*(PART 2), 313–325. https://doi.org/10.1007/978-3-540-87481-2_21

Schowe, B. (2010). Feature selection for high-dimensional data with RapidMiner. In *Technical University*. Retrieved January 10, 2020, from www-ai.cs.uni-dortmund.de/PublicPublicationFiles/schowe_2011a.pdf

Schowe, B., & Morik, K. (2011). Fast-ensembles of minimum redundancy feature selection. In *Studies in Computational Intelligence* (Vol. 373, pp. 75–95). https://doi.org/10.1007/978-3-642-22910-7_5

Sutha, K., & Tamilselvi, J. . (2015). A review of feature selection algorithms for data mining techniques. In *International Journal of Computer Science and Engineering* (Vol. 7). https://doi.org/10.1111/j.1532-849X.2011.00718.x

Tang, J., Alelyani, S., & Liu, H. (2014). *Feature selection for classification: a review BT - Data classification: algorithms and applications*.

Yang, C., Harkreader, R., & Gu, G. (2013). Empirical evaluation and new design for fighting evolving twitter spammers. In *IEEE Transactions on Information Forensics and Security* (Vol. 8). https://doi.org/10.1109/TIFS.2013.2267732

Bibliographic information of this paper for citing:

Ramzanzadeh Rostami, R., & Karbasi, S. (2020). "Detecting fake accounts on Twitter social network using multi-objective hybrid feature selection approach." *Webology*, 17(1), Article 204. Available at: http://www.webology.org/2020/v17n1/a204.pdf