IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 17, 2019, accepted February 19, 2019, date of publication March 7, 2019, date of current version March 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901756

A Spammer Identification Method for Class Imbalanced Weibo Datasets

WENBING TANG[1], ZUOHUA DING[1], (Member, IEEE), AND MENGCHU ZHOU[2], (Fellow, IEEE)

[1]School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2]Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA

Corresponding author: Zuohua Ding (zouhuading@hotmail.com)

ABSTRACT Nowadays, Weibo has become a significant and popular information sharing platform in China. Meanwhile, spammer identification has been a big challenge for it. To mitigate the damage caused by spammers, classification algorithms from machine learning have been applied to distinguish spammers and non-spammers. However, most of the previous studies overlook the class imbalance problem of real-world data. In this paper, by analyzing the characteristics of spammers in Weibo, we select microblog content similarity, the average number of links, and the other 12 features to construct a comprehensive feature vector never seen before. Considering the existence of imbalance problems in spammer identification, an ensemble learning method is used to combine multiple base classifiers for improving the learning performance. During the training stage of base learners, fuzzy-logic-based oversampling and cost-sensitive support vector machine are considered to tackle imbalanced data at both data and algorithmic levels. The experimental results demonstrate that compared with the existing state-of-the-art methods, the recall rate of our proposed approach increases by 6.5% and reaches the precision value of 87.53% when used to deal with real-world Weibo datasets we collected.

INDEX TERMS Class imbalance problem, cost-sensitive SVM, ensemble learning, fuzzy-based oversampling, spammer identification.

I. INTRODUCTION

Spamming has been a widespread challenge for online social networks. Sina Weibo, which is similar to Twitter, has become an important online social network platform in China. While providing users with real-time information dissemination, acquisition and sharing platform, Weibo also inevitably provides an opportunity for spammers. The Weibo platform itself has employed anti-spam behavior technology, such as the initial blacklist and today's user privacy permissions settings. However, with the surge in the number of Weibo users, these techniques tend to be less and less effective.

In order to improve the normal quality of online social network services and maintain cyberspace security, researchers have made some achievements in the field of spammer identification. Their studies are summarized in detail in literatures [1], [2]. These existing studies generally focus on features by mining, such as content features [3], [4] and user relationship features [5], and then the classification algorithms in machine learning, such as Random Forest [6], Decision Tree [7] are applied to distinguish between spammers and non-spammers. However, most of them overlook the class imbalance problem of real-world data [8]. Liu *et al.* [8] proved through experiments that if the Imbalance Rate (IR) rises from 2 to 20 in Twitter datasets, the spammer's detection rate drops by 33% and the non-spammer's error rate decreases by 5%, since in such scenario, a conventional classifier is biased towards the major classes (non-spammers). In fact, the proportion of spammers is far less than the proportion of non-spammers. The proportion of spammers on Twitter is about 3.75%, and Weibo is approximately 10% [9]. For such a class imbalance problem, it is difficult to obtain a satisfactory classification performance when using traditional classification methods, which has also been documented in other fields of imbalanced classification [10]–[15].

The current approaches to solve the aforementioned class imbalance problem mainly fall into two categories [16], [17] :dataset level methods, which reduces the imbalance rate

The associate editor coordinating the review of this manuscript and approving it for publication was An-An Liu.

by adjusting the class distribution, with undersampling and oversampling as their representatives; algorithm level methods, which are modified from original machine learning(data mining) algorithms for adapting the needs of real-world applications with imbalanced datasets, such as Cost-sensitive Learning, One-Class Learning. However, the work in [18] shows that simply improving from the data or algorithm level is not able to deal well with the class imbalance problem, so data and algorithms are considered at two levels to deal with the class imbalance in this work. This idea has been applied in other similar fields where imbalance exists, such as software defect prediction [19].

To solve the aforementioned class imbalance problem in Weibo spammers identification, we use a combination of Fuzzy-logic-based Oversampling(FOS) [8], Random Oversampling (ROS), Random Undersampling (RUS) and Cost-Sensitive Support Vector Machine (CS-SVM) [20] to build an ensemble classifier. At the data level, our proposed approach adopts FOS, ROS, and RUS to adjust the distribution of raw imbalanced datasets. At the algorithm level, we employ the approach of CS-SVM to deal with class imbalance by allocating different misclassification costs, which allows us to specify the relative importance of different kinds of prediction errors. The proposed approach consists of three steps. The first step builds three different classes of balanced training datasets, which is obtained by adjusting the distribution of the original imbalanced training dataset by FOS, ROS, and RUS. The second one aims to train a classifier from each training dataset obtained in the previous step by CS-SVM. The last one combines the base classifier results to obtain the final classification results by majority voting scheme.

In this paper, we firstly analyze the characteristics of spammers in Weibo and then select Weibo content similarity, the average number of links and other twelve features to construct a comprehensive feature vector never seen before. Secondly, we propose an ensemble method for distinguishing Weibo spammers from non-spammers on class-imbalanced data. We then compare our proposed method with the existing similar methods. In particular, we compare the performance of different methods under different class imbalance rates. The experimental results show that our proposed approach is an efficient identification method in Weibo spammer detection when the data are imbalanced. For example, when the class imbalance rate equals 10 in the dataset, the recall rate of our learning approach increases by 6.5% compared with the existing state-of-the-art methods. The major contributions of this paper are given as follows:

1) We analyze the characteristics of spammers in Weibo, and then select Weibo content similarity, the average number of links, and other twelve features to construct a comprehensive feature vector never seen before.
2) In order to address the class imbalance problem in Weibo, we develop an ensemble identification method, which deals with imbalance problem from both data and algorithm levels.
3) Multiple comprehensive comparative studies are conducted based on our proposed method and other similar models for Weibo spammer identification, and the experimental results show the effectiveness and robustness of our approach in class imbalance Weibo dataset.

The rest of this paper is organized as follows. Section II summarizes related work. Section III introduces the general background of Weibo spammers identification. We further analyze the spammer's characteristics and then explains the spammer features used in this paper. In Section IV, we interpret the proposed ensemble method for spammer identification based on FOS and CS-SVM. The comparative experimental results and analysis are given in Section V, conclusions and future research directions are finally provided in Section VI.

II. RELATED WORK

This section presents related work on two subareas: spammer detection approaches and class imbalance problem about spammer detection.

A. SPAMMER DETECTION APPROACHES

In order to detect Weibo(Twitter) spammers, there are many studies have been published. Most of these works are using machine learning algorithm.

For example, McCord and Chuah [6] applied traditional machine learning classifiers such as Random Forest, Naive Bayesian, SVM and k-NearestNeighbor(k-NN) on Twitter data they collected. Their results showed that the Random Forest algorithm has the best performance. Wang [21] indicated that Naive Bayesian has better classification results than the Neural Network, Decision Tree, SVM and k-NN on the dataset he collected.

Zheng *et al.* [22] proposed an Extreme Learning Machine-based classification algorithm for spammer detection on the Sina Weibo dataset they collected. In their proposed approach, through crawling Sina Weibo data and labeling corresponding samples into spammers and non-spammers manually. Meda *et al.* [23] combined feature selection with Random Forest algorithm to detect spammers in Twitter. Later, they validated the effectiveness of the Random Forest algorithm compared with the Support Vector Machine and the Extreme Learning Machines [3].

In addition, Wu *et al.* [24], tried to use deep learning methods in Twitter spammer detection. In their proposed method, Word2Vec is applied to pre-process the tweets. Mukherjee *et al.* [25] proposed an unsupervised Bayesian inference framework to detect opinion spammers(i.e., fake reviewers). Different from most existing supervised or unsupervised learning models, Wu *et al.* [26], [27] suggested that semi-supervised learning method can be applied in spammer detection. They considered that there may be only a small amount of labeled data and a large amount of unlabeled data in many scenarios.

At present, the features used to distinguish between spammers and non-spammers mainly include account-based features [3], content-based features [24], relationship(graph)-based features [5], [28], behavior-based features [25], [27], time-based features [29], and hybrid features [1], [2].

For example, Hu *et al.* [30] attempted to use network and content information to spammer detection collectively. Liu *et al.* [31] employed user behavior information, online social network attributes, and text content characteristics as features for spammer detection. Vishwarupe *et al.* [32] used the type of account as a feature, i.e., they checked whether the account was verified.

Different from these existing work, we select twelve features related to content, relationship, behavior, device, and time to construct a feature vector. Furthermore, when labeling a dataset record, we check that the user's account state is verified or not.

B. CLASS IMBALANCE PROBLEM IN SPAMMER DETECTION

There are often very few spammers' samples compared to a large number of non-spammer samples [33], In such data imbalance scenario, it is difficult to obtain a satisfactory classification performance when using traditional classification. Therefore, a key question in Weibo spammers identification is how to improve the performance of classifier(identification method) facing with class imbalance data.

In the past, it was often random to select an approximately equal number of positive and negative class samples from the data set, e.g., the work in [23]. However, the training set constructed in this way is not consistent with the class distribution of real-world samples.

In order to make the class distribution of training dataset more consistent with real-world, some researchers have paid attention to the class imbalance problem of spammer identification. Jin *et al.* [9] proposed an ensemble Undersampling-based strategy to tackle the class imbalance problem on spam detection in Weibo. A novel way named fuzzy-logic-based oversampling (FOS) [34] has been proposed to achieve balanced class distribution through a fuzzy-logic-based information decomposition algorithm [35]. Later, Liu *et al.* [8] first investigated the class imbalance problem in the domain of Twitter spammers detection, and then combine FOS with ensemble learning to handle the class imbalance problem in Twitter datasets.

III. FEATURE ANALYSIS AND DEFINITIONS

In this section, we introduce the general background of Weibo spammers identification and then discuss the features used in this work for distinguishing spammers, i.e., how to construct a user's feature vector.

A. THE GENERAL BACKGROUND OF WEIBO SPAMMERS IDENTIFICATION

The spammers identification can be viewed as a binary classification problem, i.e., a mapping f from the input space(feature space) $\mathcal{X}$ to the output space $\mathcal{Y}$ is established: $\mathcal{X} \to \mathcal{Y} = \{spammer, non\_spammer\}$. Formally describe it as follows.

$$\{features \quad set\}_u \xrightarrow{f} \{spammer, non\_spammer\}. \tag{1}$$

Based on the given dataset, we can train a classifier using our proposed learning algorithm, which can predict whether any given testing user labels to spammer or non-spammer.

Before giving a solution for identifying spammers, we first define the user's features, which will be used later.

TABLE 1. Weibo users feature set.

No.	Dimension	Feature	Description
1	Content	Similarity	The average similarity of Weibo content
2		Ave_URL	The average number of links included in each Weibo
3		Ave_tran	The number of times each Weibo was reposted
4		Ave_comm	The average amount of comments on each Weibo
5		Ave_digit	The average number of digits contained in each Weibo
6		Ave_symbol	The average number of symbols included in each Weibo
7	Relationship	Reputation	Used to measure the user's attention
8		Mutual_rate	Percentage of mutual friends
9	Behavior	Tran_rate	The proportion of reposted microblog in all Weibo
10		Ave_ment	The average number of "@"
11	Device	PC_rate	Measure the percentage of users use Weibo via PC
12	Time	Min2	Number of posts per minute over 2 times

B. DEFINITIONS OF USER FEATURES

We all know that spammers identification is based on a range of features. Hence, in this part, we discuss the features used in this paper.

By analyzing the characteristics of spammers, we consider five-dimensional features: content, relationship, behavior, device, and time. Table 1 shows the description of these features.

1) CONTENT-BASED FEATURES

The content-based features mean properties of the text of microblog posted by users. As we all know, spammers prefer to include links, numbers and symbols in their microblogs, and they like to transmit other microblogs in large quantities. In this way, they achieve the purpose of fraud, harassment and so on.

Here, we define the features:*similarity*, *ave_URL*, *ave_tran*, *ave_comm*, *ave_digit*, and *ave_symbol* as content-based features. The detailed definitions of these features are as follows.

Similarity: The similarity of microblog contents is used to measure a user's average similarity and repeatability of microblog contents. The *similarity* of non-spammers tend to be smaller than that of spammers. For a user u, $sim(i, j)$ is

the content similarity between i-th and j-th microblogs of u. Then:

$$Similarity = \frac{2}{a(a-1)} \sum_{i=1}^{a} \sum_{j=i+1}^{a} Sim(i, j). \quad (2)$$

where a indicates the total number of microblogs of user u. The $sim(i, j)$ is calculated by using the short text similarity function of Baidu AI open platform.[1]

Plainly, many spammers' microblogs may contain malicious URL(Uniform Resource Locator) to mislead normal users to click it. *Ave_URL* is definded to measure the average number of URL included in each Weibo.

$$Ave\_URL = \frac{No\_URL}{a}. \quad (3)$$

where, *No_URL* represents the number of links for user u in an among all microblogs. The definitions of *ave_tran*, *ave_comm*,*ave_digit* and *ave_symbol* are similar to *ave_URL*.

2) RELATIONSHIP-BASED FEATURES

The relationship-based features are the features involving the Weibo users' social relationship, including the number of following, followers and friends. It was obvious that spammers usually have a large number of fans in order to spread malicious information and attract more people's attention, while non-spammers use Weibo to browse interesting contents. Hence, the number of attention is often greater than the number of fans. *Reputation*(*attention*), used to measure the user's social network characteristics.

$$Reputation = \frac{No\_fan}{No\_att + No\_fan} \quad (4)$$

where *no_fan* represents the number of fans of user u, and *no_att* indicates the number of attentions. However, *reputation* only considers unilateral relationships and does not take into account the bidirectional relationship. For this purpose, the *mutual_rate* is defined to further analyze user relationships.

$$Mutual\_rate = \frac{No\_mut}{No\_att + No\_fans - No\_mut} \quad (5)$$

where *no_mut* represents the number of mutual followers (friends). The non-spammer's *mutual_rate* tend to be higher.

3) BEHAVIOR FEATURES

Spammers often repeatly posts a microblog to attract attention within a short period of time, or frequently "@" other users to disseminate malicious information. We consider the following metrics as user behavior features: The proportion of reposted microblog in all Weibo(i.e., *tran_rate*), The average number of "@" in all Weibo text, *ave_ment*. The definitions of *tran_rate* and *ave_ment* are similar to *ave_URL*.

[1] http://ai.baidu.com/tech/nlp/simnet

4) DEVICE FEATURES

Spammers prefer using mobile devices to computers. Hence, we define *PC_rate* to measure the percentage of users posted Weibo via PC. The non-spammer's *PC_rate* tends to be higher.

5) TIME FEATURES

Since writing Weibo requires editing and typing, non-spammers and spammers exhibit different statistical characteristics at the time of posting Weibo. For this reason, we define the number of times that microblogs are posted more than twice (threshold) per minute, expressed as *min*2.

IV. PROPOSED ENSEMBLE IDENTIFICATION METHOD

In this section, we introduce our proposed novel ensemble spammer identification method. Our method uses a combination of FOS, ROS, RUS and CS-SVM to build an ensemble classifier to deal with imbalanced data. Now suppose we are given a dataset consisting of n spammers and m non-spammers, where $n \ll m$. That is,

$$\begin{aligned} D &= D_+ \cup D_- \\ &= \{(x_1, y_+), \ldots, (x_n, y_+), (x_{n+1}, y_-), \ldots, (x_{n+m}, y_-)\} \end{aligned} \quad (6)$$

As shown in Fig. 1, the proposed ensemble method consists of three stages. The three stages are described in detail below.

A. FEATURE EXTRACTION STAGE

The main purpose of this stage is to extract user features for identification. For each user, 12 features are extracted. The feature set of i-th user can be expressed as x_i, and:

$$x_i = \{x_{i1}, \cdots, x_{i12}\} \quad (7)$$

where, from x_{i1} to x_{i12} represent 12 features of a user. which are calculated according to the description in Section III.

B. TRAINING SET REBALANCE STAGE

At this stage, the original training set is re-sampled by FOS, ROS and RUS methods and then we get three different balanced training datasets.

1) FUZZY-BASED OVERSAMPLING (FOS)

When the number of positive examples is n, the oversampling rate α is given. The basic idea of FOS [34] is: divide each feature dimension into $q = n \times \alpha$ feature subintervals. Secondly, the membership function u is used to measure the membership degree (i.e., weight) of each observed value on each feature subinterval. Finally, these weights are used to generate data in each feature dimension. The details of FOS are as follows:

The 1st Step (Partition of Feature Subintervals): It is assumed that under the current feature dimension, the observed value $O_i = \{x_{1i}, x_{2i}, \cdots, x_{ni}\}$. Assuming that *min* and *max* are the minimum and maximum values of

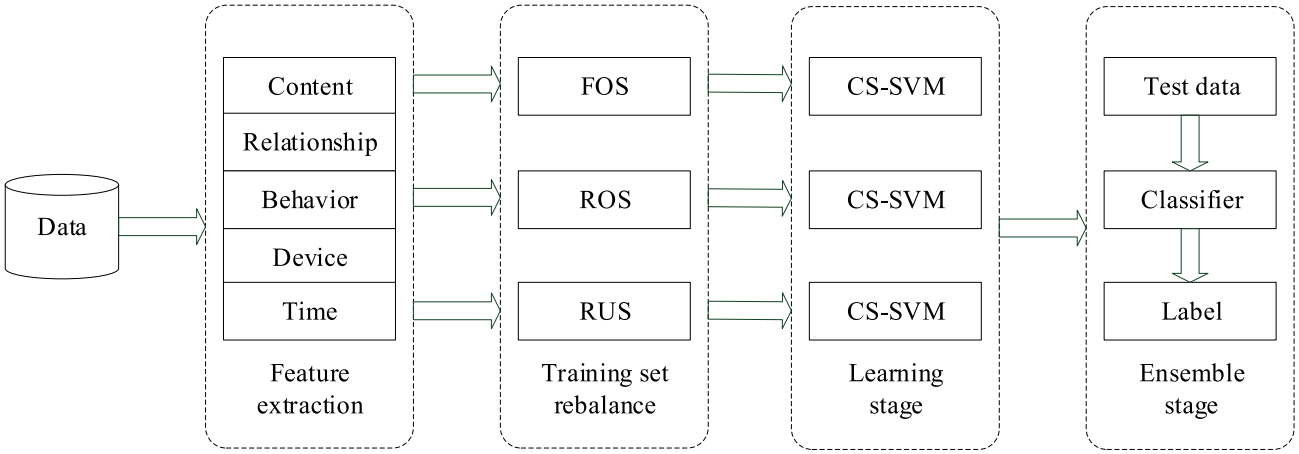

FIGURE 1. Framework of the proposed ensemble identification method.

observations respectively, i.e.,

$$x_j^m = \min_j\{x_{ji}\}$$
$$x_j^M = \max_j\{x_{ji}\} \tag{8}$$

FOS divides the interval $[x_j^m, x_j^M]$ equidistant into q subintervals, h represents the length of the subinterval, and $c_s(s = 1, 2, \ldots, q)$ represents the center of each subinterval.

The 2nd Step (Calculation of Membership Degree): Each $x_{ji}(j = 1, 2, ..., n)$ has q membership functions. Let $\mu(x_{ji}, c_{js})$ denote the s-th membership function of x_{ji}.

$$\mu(x_{ji}, c_{js}) = \begin{cases} 1 - \dfrac{\| x_{ji} - c_{js} \|}{h}, & \| x_{ji} - c_{js} \| \leq h \\ 0, & otherwise \end{cases} \tag{9}$$

The 3rd Step (Generating Synthetic Data): Based on the membership function, new synthetic data can be generated. If all observation values' membership function is zero on a subinterval, then the average value of all observed values is used instead of the generated data in this interval. Otherwise, the weighted mean is set as the generated data value. Here is the rule of data generation.

$$\tilde{O_{js}} = \begin{cases} \bar{x_i}, & \sum_{i=1}^{n} m(x_{ji}, c_{js}) = 0 \\ \dfrac{\sum_{i=1}^{n} m(x_{ji}, c_{js}) \times x_{ji}}{\sum_{i=1}^{n} m(x_{ji}, c_{js})}, & otherwise \end{cases} \tag{10}$$

Finally, a new training set D' is obtained by combining the positive dataset obtained from FOS with the original negative dataset D_-, i.e., $D' = D'_+ \cup D_-$.

2) RANDOM OVERSAMPLING (ROS)

ROS is a basic oversampling method, which could help to achieve balance class distribution by replicating minority class sample. However, ROS may cause the model to be overfitted, which also increases the size of the training set [11].

3) RANDOM UNDERSAMPLING (RUS)

RUS tries to balance class distribution through the random elimination of majority class examples. The issue of RUS is that it may remove some useful information [33].

In order to construct a class balanced training set, in this paper, we combine these three sampling strategies with different advantages and disadvantages to generate a training set with approximate balance class distribution.

C. LEARNING STAGE

At this stage, CS-SVM [20] is used for learning on each balanced training dataset obtained from training set rebalance stage. In the next part, we will introduce the CS-SVM used in this paper.

1) COST-SENSITIVE SUPPORT VECTOR MACHINE (CS-SVM)

CS-SVM is an improvement of classic C-SVM (Soft-margin SVM) proposed by Vapnik [37] to achieve cost sensitivity for addressing a class imbalance problem. C-SVM assigns the same misclassification cost for both positive and negative examples when it encounters imbalanced data. It leads to a separating hyper plane biased towards the major class.

To solve the aforementioned problem, Veropoulos *et al.* [20] suggest using different misclassification costs for positive and negative classes, making SVM applicable to class imbalance problems.

The CS-SVM optimization problem is:

$$\begin{aligned} \min_{w,b,\xi} \ & \frac{1}{2} \| w \|^2 + C^+ \sum_{\{i|y_i=+1\}}^{n+m} \xi_i + C^- \sum_{\{i|y_i=-1\}}^{n+m} \xi_i \\ s.t. \ & y_i(w \cdot \Phi(x_i) + b) \geq 1 - \xi_i \\ & xi_i \geq 0, \quad i = 1, 2, \cdots, n+m \end{aligned} \tag{11}$$

where $w \cdot \Phi(x_i) + b = 0$ indicates the separating hyper plane, xi_i is a slack variable. C^+ and C^-represent misclassification costs of positive and negative classes, respectively. CS-SVM can mitigate the effects of class imbalance by setting higher

misclassification costs (i.e., $C^+ > C^-$) to positive class than negative class.

The Lagrange function obtained by using the Lagrangian multiplier method is:

$$L(w, b, \xi, \alpha, \gamma) = \frac{1}{2} \parallel w \parallel^2 + C^+ \sum_{\{i|y_i=+1\}}^{n+m} \xi_i + C^- \sum_{\{i|y_i=-1\}}^{n+m} \xi_i - \sum_{i=1}^{n+m} \alpha_i [y_i (w \cdot \Phi(x_i) + b) - 1 + x_i] - \sum_{i=1}^{n+m} \gamma_i x_i \quad (12)$$

where $\alpha_i \geq 0$ and $\gamma_i \geq 0$ are Lagrange multipliers. Letting L's partial derivative with respect to w, b and ξ be 0, we can get the following dual problem:

$$\max_{\alpha} \sum_{i=1}^{n+m} \alpha_i - \frac{1}{2} \sum_{i=1}^{n+m} \sum_{j=1}^{n+m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$s.t. \sum_{i=1}^{n+m} \alpha_i y_i = 0, \quad 0 \leq \alpha_i^+ \leq C^+, \ 0 \leq \alpha_i^- \leq C^-$$
$$i = 1, 2, \cdots, n+m \quad (13)$$

where α_i^+ and α_i^- are Lagrange multipliers for positive and negative cases, respectively. Solving this dual problem, we can get α, and then find w and b, which is the solution to the original problem.

In order to improve the diversity of the base learner [38], three basic CS-SVM learners use the Gaussian, polynomial and Sigmoid kernel function, respectively.

D. ENSEMBLE STAGE

In the last stage, Considering that ensemble is often better than the best single [39], we combine the identification results of three basic learners obtained from the above stage. The majority voting method was selected as the combined strategy, which is the most popular voting method.

The proposed ensemble identification algorithm is illustrated in Algorithm 1.

Algorithm 1 The Learning Algotithm of Our Proposed Method

Setp-1 Data Sampling
Input: Training dataset $D(D = D_+ \cup D_-)$, Oversampling ratio α,Undersampling ratio β.
Output: Three different balanced training set D_1,D_2 and D_3.
$D_1 = \text{FOS}(D_+, \alpha) + D_-$;
$D_2 = \text{ROS}(D_+, \alpha) + D_-$;
$D_3 = \text{RUS}(D_-, \beta) + D_+$;
Setp-2 Training
Input: Balanced training set D_1,D_2, and D_3.
Output: Multiple base classifiers C_1,C_2 and C_3.
Train classifier C_1 using CS-SVM algorithm with Gaussian kernel function in D_1;
Train classifier C_2 using CS-SVM algorithm with polynomial kernel function in D_2;
Train classifier C_3 using CS-SVM algorithm with Sigmoid kernel function in D_3;
Setp-3 Ensemble
Input: Tast sample x.
Output: predicted class of x.
Initialization: $Z\_spammer = 0$,$Z\_non\text{-}spammer = 0$.
for $i = 1; i <= 3; i++$ **do**
  if $C_i(x) = spammer$ **then**
    $Z\_spammer = Z\_spammer + 1$;
  else
    $Z\_non\text{-}spammer = Z\_non\text{-}spammer + 1$;
  end if
end for
if $Z\_spammer > Z\_non\text{-}spammer$ **then**
  return *spammer*
else
  return *non-spammer*
end if

V. EXPERIMENTS

A. DATASET COLLECTION

Although Weibo has opened more than 20 types of interfaces including microblog, comments, users, and relationships, there are still many restrictions, such as limiting the frequency of access. Therefore, in this paper, we collect data through web crawlers written in python.

Firstly, we randomly selected 200 verified users from Weibo social network. Considering that most ordinary users are unlikely to follow spammers [22], we crawl the follow of these users. Therefore, we can get a lot of non-spammer information according to this approach.

For spammers, we first obtained some spammer IDs from the Sina Weibo Community Management Center,[2] which is a platform to deal with violations of the legitimate rights and interests of users [31]. In addition, we got some spammers' IDs by using the function of finding people based on keywords, this way has also been applied in [9]. After that, we crawled their feature information based on these IDs. Their followers are also considered spammers, and then climb the features of these followers.

Through the aforementioned two approaches, We collect the latest 200 microblogs for each user. If the number of microblogs is less than 200, we collect all the microblogs for that user. Finally, we can get the dataset needed in this paper, which contains a total of 17,115 user records. The IR values on a training and test set are 9.93 and 10.01, respectively. A statistic summary of two datasets are shown in Table 2:

[2]http://service.account.weibo.com

TABLE 2. Statistics of training and test dataset.

Dataset	# of spammer	# f non-spemmer	Total
Training set	1208	11998	13206
Test set	355	3554	3909

TABLE 3. Confusion matrix.

Human Classification	Prediction	
	Spammer	Non-spammer
Spammer	TP	FN
Non-Spammer	FP	TN

B. PARAMETERS SETTING

Before starting experiments, we need to set the parameters of the method. In the training set, in order to make IR less than 4 [40], we choose $\alpha = 3$. In practice, the parameters in CS-SVM are usually set as follows:

$$C^+ = \frac{m}{m+n}C$$
$$C^- = \frac{n}{m+n}C \tag{14}$$

where we select 200 as the value of C [41].

C. EVALUATION METRICS

For a spammer identification method, according to the combination of the actual class and the classification result, the confusion matrix is depicted in Table 3.

Based on the above confusion matrix, we use the following metrics to measure the identification performance, namely, *precision*, *recall*, and $F1$-*score*.

1) PRECISION

Precision indicates the ratio of correctly predicted positive samples to the total predicted positive samples.

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

2) RECALL

Recall calculates the proportion of all ''correctly identified samples'' to all ''samples that should be identified''.

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

3) F1-SCORE

$F1$-*score* is the weighted average of Precision and Recall. The formula for the $F1$-*score* is:

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

D. EXPERIMENTAL RESULTS AND ANALYSIS

The main purpose of experiments is to use the method we proposed in this paper to judge whether a Weibo user is a spammer or non-spammer, and compare our method with the existing spammer identification methods on imbalanced datasets for Twitter and Weibo platforms, so as to illustrate the effectiveness and robustness of our approach in dealing with the class imbalance datasets.

TABLE 4. Identification performance on imbalanced dataset with the IR equals 10.

	Random Forest	RUSBoost [9]	FOSEnsemble [8]	Our Approach
$Precision$	0.808	0.781	**0.916**	0.875
$Recall$	0.527	0.698	0.835	**0.9**
$F1-score$	0.638	0.737	0.873	**0.887**

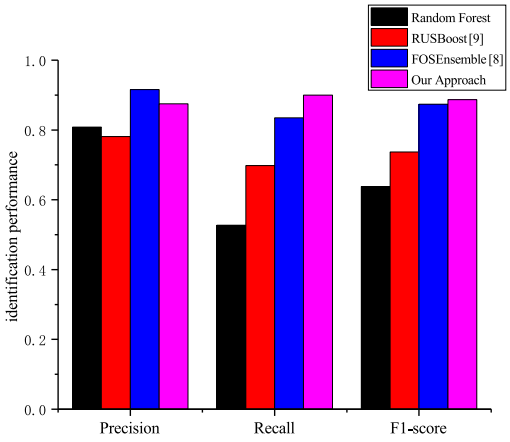

FIGURE 2. The performance of our proposed method are compared with existing methods on imbalanced dataset when IR equals 10.

To evaluate the spammer identification performance, we compare our approach with the spammer identification techniques for imbalanced datasets in [8] and [9], and the state-of-the-art random forest method for balanced datasets with the highest accuracy [2]. On the same dataset, the comparative experimental results are shown in Table 4. We independently conducted 20 experiments to take the average value as the final experimental results. Table 4 shows that our approaches *recall* value (90%) is higher than those of other three methods, and it achieves a lower $F1$-*score* 88.7%.

Fig. 2 presents the identification performance of each method in dimensions of *precision*, *recall*, and $F1$ -*score*. Although the random forest method can achieve very good (up to 95.7%) classification results on class balanced datasets [6], when applied to the problem of Weibo spammer identification with class imbalance the *recall* value is very low. RUSBoost [9] can significantly enhance the *recall* value compared to the random forest method, but its *precision* is not well. FOSEnsemble [8] has a better *precision* and *recall* compared with Random Forest and RUSBoost. Although our approach is similar to FOS in terms of *precision*, we have a highest *recall*, which is important metric in the identification problem. At the same time, we can see that the $F1$-*score* and *recall* value of our proposed method are the highest among the four methods. All these show that our proposed method is an effective and reliable way to handle the imbalances dataset we collected.

TABLE 5. F1-score in dataset with varying class imbalance rate.

IR	Random Forest	RUSBoost [9]	FOSEnsemble [8]	Our Approach	Average value
2	**0.865**	0.787	0.854	0.849	0.838
5	0.715	0.742	0.861	**0.878**	0.799
10	0.638	0.737	0.874	**0.887**	0.784

In addition, we also compare the $F1$-*score* of the above four methods under different class imbalance rates, i.e., IR equals 2, 5, 10, respectively. The results are depicted in Table 5. When IR = 2, the $F1$-*score* of our proposed methods is 0.849, and the random forest is the highest, reaching 0.865. However, as the imbalance rate increases, there is no large fluctuation in $F1$-*score* with our method. The $F1$-*score* of random forest decreased sharply with the increase of IR. As can be seen from Table 5, the $F1$-*score* of our proposed method is higher than that of the other three methods when the dataset is imbalanced. Table 5 indicates the robustness of our approach in Weibo spammer identification.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a Weibo spammer identification method based on FOS and CS-SVM ensemble. Based on the analysis of spammers features, we construct a feature vector. We used ensemble learning methods to improve the performance of base learners. Considering the existence of imbalance problems in spammer identification, FOS and CS-SVM are used to deal with imbalanced data in the training process of the base learner in both data and algorithm levels. Through comprehensive analysis and comparison, the effectiveness of our method on the collected datasets is better than other existing methods, i.e., our method can more effectively differentiate spammers from non-spammers. However, due to the features used in this paper, such as microblog content similarity, more complex calculations are required.

Our future work will focus on two aspects: firstly, in the process of oversampling, in order to generate representative samples, we can consider to generate data by using generative adversarial networks (GAN). Secondly, we will combine the proposed method with transfer learning to deal with similar problems in other platforms.

REFERENCES

[1] A. T. Kabakus and R. Kara, "A survey of spam detection methods on Twitter," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, pp. 29–38, 2017.
[2] P. Kaur, A. Singhal, and J. Kaur, "Spam detection on Twitter: A survey," in *Proc. INDIACom*, New Delhi, India, Mar. 2016, pp. 2570–2573.
[3] C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, "Machine learning techniques applied to Twitter spammers detection," in *Proc. ICCST*, Rome Italy, 2014, pp. 177–182.
[4] C. Lin, J. He, Y. Zhou, X. Yang, K. Chen, and L. Song, "Analysis and identification of spamming behaviors in sina weibo microblog," in *Proc. 7th Workshop Social Netw. Mining Anal.*, New York, NY, USA, 2013, p. 5.
[5] J. Song, S. Lee, and J. Kim, "Spam filtering in Twitter using sender-receiver relationship," in *Proc. RAID*, Sacramento, CA, USA, 2011, pp. 301–317.
[6] M. Mccord and M. Chuah, "Spam detection on Twitter using traditional classifiers," in *Proc. Int. Conf. Autonomic Trusted Comput.*, Edmonton, AB, Canada, 2011, pp. 175–186.
[7] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. N. Choudhary, "Towards online spam filtering in social networks," in *Proc. NDSS*, Sacramento, CA, USA, 2012, pp. 1–16.
[8] S. Liu, Y. Wang, J. Zhang, C. Chen, and Y. Xiang, "Addressing the class imbalance problem in Twitter spam detection using ensemble learning," *Comput. Secur.*, vol. 69, pp. 35–49, Aug. 2017.
[9] Z. Jin, Q. Li, D. Zeng, and L. Wang, "Filtering spam in Weibo using ensemble imbalanced classification and knowledge expansion," in *Proc. ISI*, Baltimore, MD, USA, May 2015, pp. 132–134.
[10] H. Liu, M. Zhou, X. S. Lu, and C. Yao, "Weighted Gini index feature selection method for imbalanced data," in *Proc. ICNSC*, Zhuhai, China, Mar. 2018, pp. 1–6.
[11] Q. Kang, X. Chen, S. Li, and M. Zhou, "A noise-filtered under-sampling scheme for imbalanced classification," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4263–4274, Dec. 2017.
[12] F. Wang, T. Xu, T. Tang, M. Zhou, and H. Wang, "Bilevel feature extraction-based text mining for fault diagnosis of railway systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 49–58, Jan. 2017.
[13] S. Teng, N. Wu, H. Zhu, L. Teng, and W. Zhang, "SVM-DT-based adaptive and collaborative intrusion detection," *IEEE/CAA J. Autom. Sin.*, vol. 5, no. 1, pp. 108–118, Jan. 2018.
[14] P. Zhang, S. Shu, and M. Zhou, "An online fault detection model and strategies based on svm-grid in clouds," *IEEE/CAA J. Autom Sin*, vol. 5, no. 2, pp. 445–456, Mar. 2018.
[15] W. Zhang, H. Zhang, J. Liu, K. Li, D. Yang, and H. Tian, "Weather prediction with multiclass support vector machines in the fault detection of photovoltaic system," *IEEE/CAA J. Autom. Sin.*, vol. 4, no. 3, pp. 520–525, Jul. 2017.
[16] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, "On the class imbalance problem," in *Proc. ICNC*, Jinan, China, Oct. 2008, pp. 192–201.
[17] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[18] V. Effendy, Adiwijaya, and Z. K. A. Baizal, "Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest," in *Proc. ICoICT*, Bandung, Indonesia, May 2014, pp. 325–330.
[19] S. Huda *et al.*, "An ensemble oversampling model for class imbalance problem in software defect prediction," *IEEE Access*, vol. 6, pp. 24184–24195, 2018.
[20] K. Veropoulos, C. Campbell, and N. Cristianini, "Controlling the sensitivity of support vector machines," in *Proc. IJCAI*, Stockholm, Sweden, 1999, p. 60.
[21] A. H. Wang, "Machine learning for the detection of spam in Twitter networks," in *Proc. ICETE*, Athens, Greece, 2010, pp. 319–333.
[22] X. Zheng, X. Zhang, Y. Yu, T. Kechadi, and C. Rong, "ELM-based spammer detection in social networks," *J. Supercomput.*, vol. 72, no. 8, pp. 2991–3005, 2016.
[23] C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, "A machine learning approach for Twitter spammers detection," in *Proc. ICCST*, Rome, Italy, Oct. 2014, pp. 1–6.
[24] T. Wu, S. Liu, J. Zhang, and Y. Xiang, "Twitter spam detection based on deep learning," in *Proc. ACSW*, Geelong, Australia, 2017, p. 3.
[25] A. Mukherjee *et al.*, "Spotting opinion spammers using behavioral footprints," in *Proc. ACM SIGKDD*, Chicago, IL, USA, 2013, pp. 632–640.
[26] Z. Wu, J. Wu, J. Cao, and D. Tao, "HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation," in *Proc. ACM SIGKDD*, Beijing, China, 2012, pp. 985–993.
[27] Z. Wu, J. Cao, Y. Wang, Y. Wang, L. Zhang, and J. Wu, "hPSD: A hybrid PU-learning-based spammer detection model for product reviews," *IEEE Trans. Cybern.*, to be published. doi: 10.1109/TCYB.2018.2877161.
[28] X. Wu, Z. Feng, W. Fan, J. Gao, and Y. Yu, "Detecting marionette microblog users for improved information credibility," in *Proc. IJCAI*, Beijing, China, 2013, pp. 483–498.
[29] H. Liu, Y. Zhang, H. Lin, J. Wu, Z. Wu, and X. Zhang, "How many zombies around you?" in *Proc. IEEE ICDM*, Dallas, TX, USA, Dec. 2013, pp. 1133–1138.
[30] X. Hu, J. Tang, Y. Zhang, and H. Liu, "Social spammer detection in microblogging," in *Proc. IJCAI*, Beijing, China, 2013, pp. 2633–2639.
[31] Y. Liu, B. Wu, B. Wang, and G. Li, "SDHM: A hybrid model for spammer detection in Weibo," in *Proc. ASONAM*, Beijing, China, Aug. 2014, pp. 942–947.
[32] V. Vishwarupe, M. Bedekar, M. Pande, and A. Hiwale, "Intelligent Twitter spam detection: A hybrid approach," in *Proc. Smart Trends Syst., Secur. Sustainability*, London, U.K., 2018, pp. 189–197.

[33] S. Liu, J. Zhang, Y. Xiang, and W. Zhou, "Fuzzy-based information decomposition for incomplete and imbalanced data learning," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1476–1490, Dec. 2017.
[34] S. Liu, Y. Wang, C. Chen, and Y. Xiang, "An ensemble learning approach for addressing the class imbalance problem in Twitter spam detection," in *Proc. ACISP*, Melbourne, Australia, 2016, pp. 215–228.
[35] S. Liu, J. Zhang, Y. Wang, and Y. Xiang, "Fuzzy-based feature and instance recovery," in *Proc. ACIIDS*, Da Nang, Vietnam, 2016, pp. 605–615.
[36] C. Li and S. Liu, "A comparative study of the class imbalance problem in Twitter spam detection," *Concurrency Comput., Pract. Exper.*, vol. 30, no. 5, 2017, Art. no. e4281.
[37] V. Vapnik, "The nature of statistical learning theory," in *Proc. Conf. Artif. Intell.*, 1995, pp. 988–999.
[38] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.
[39] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.
[40] E. J. R. Silva and C. Zanchettin, "On the existence of a threshold in class imbalance problems," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Hong Kong, Oct. 2015, pp. 2714–2719.
[41] Y. Wei-Yun, Q. Zheng, Z. Yu, L. Bing, and L. Xiu, "Support vector machine and its application in customer churn prediction," *Syst. Eng.-Theory Pract.*, vol. 27, no. 7, pp. 105–110, Jul. 2007.

WENBING TANG received the B.Eng. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2017. He is currently pursuing the M.S. degree with the School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China. His current research interests include class imbalance learning and intelligent control.

ZUOHUA DING (M'11) received the M.S. degree in computer science and the Ph.D. degree in mathematics from the University of South Florida, Tampa, FL, USA, in 1996 and 1998, respectively. He is currently a Professor and the Director with the Laboratory of Intelligent Computing and Software Engineering, Zhejiang Sci-Tech University, Hangzhou, China. He has authored and co-authored over 70 papers. His current research interests include system modeling, software reliability prediction, intelligent software systems, and service robots.

MENGCHU ZHOU (S'88–M'90–SM'93–F'03) received the B.S. degree in control engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1983, the M.S. degree in automatic control from the Beijing Institute of Technology, Beijing, China, in 1986, and the Ph.D. degree in computer and systems engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

He joined the New Jersey Institute of Technology (NJIT), Newark, NJ, USA, in 1990, where he is currently a Distinguished Professor of electrical and computer engineering. He has over 680 publications, including 12 books, over 360 journal papers (over 260 in the IEEE TRANSACTIONS), and 28 book chapters. His current research interests include Petri nets, sensor networks, Web services, big data, semiconductor manufacturing, transportation, and energy systems.

Prof. Zhou was a recipient of the Perlis Research Award and the Fenster Innovation in Engineering Education Award from NJIT, the Humboldt Research Award for U.S. Senior Scientists, the Leadership Award, the Academic Achievement Award from the Chinese Association for Science and Technology, USA, the Outstanding Contributions Award, the Distinguished Lecturership, the Franklin V. Taylor Memorial Award, the Norbert Wiener Award of the IEEE SMC Society, and the Distinguished Service Award from the IEEE Robotics and Automation Society. He is the Founding Editor of the IEEE Press Book Series on Systems Science and Engineering. He is a Life Member of the Chinese Association for Science and Technology, USA, and served as its President, in 1999. He is a Fellow of IFAC and AAAS.