Crank Call Detection Models Based on Call Data of Telephone Subscribers

Wenbo Xie\textsuperscript{1,2,3,*}, Zhen Liu\textsuperscript{1,2} and Yan Fu\textsuperscript{1,2}

\textsuperscript{1}Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, People’s Republic of China.
\textsuperscript{2}Web Sciences Center, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, People’s Republic of China.
Email: xwb0211@qq.com
\textsuperscript{*}Corresponding Author Email: 201611060103@std.uestc.edu

Abstract. Data explosion makes our life much more convenient than ever. For example, people always get timely news from Apps, and get interesting recommendations when shopping online. However, these conveniences come by probing people’s private data, such as account information, search history, and chat records. Thus, if the management of data is not as secure as it supposes to be, many problems, e.g., crank calls, will arise and cause big troubles in people’s lives. Although many security Apps have been developed to help mobile phone users stay away from crank calls, the users need to update Apps frequently due to their user-labels-based blacklist. These feedback-based systems depend on a large number of user actions that have significant lag. In this paper, we propose two crank call detection models to help operators mark crank callers based on the call data of telephone subscribers, in which the features are organized by different ways. One is a composited classification model, and the other is a convolutional neural network model based on the time-sliced data. The experiments show that the model applying convolutional neural network gets better accuracies in detecting crank calls.

1. Introduction
In big data area, the collection of information happens anywhere and anytime, which makes the data increasing crazily. The data explosion has brought a lot of convenience to people. For example, thanks for the user-record-based recommendation system, people can easily achieve product news they are interested in without searching in the Apps, and focus on exactly what they need to buy when they are shopping online. However, every coin has two sides, the failure of managing the excessive amount of data has also brought troubles for people, e.g., crank calls. There are many types of crank calls, in which the sales promotion is one of the most common type. According to the analysis report about China’s crank call in 2016 [1], the App CooTek has intercepted crank calls more than 32.2 billion times from August 2015 to July 2016, and there are about 90 million crank calls per day on average. In 2018, the number of crank calls in China has exceeded 50 billion, and about 121 million crank callers have been marked by 360 Mobile Phone Guards [2].

To protect telephone subscribers from crank calls, large number of security Apps have been developed, in which user-labels based collaborative filtering is the most common strategy to defeat crank calls. But this strategy has three obvious drawbacks. The first one is the limitation of users’ feedback. Different App can only get the feedback from their own users and not all the users who want to submit feedback so as to it takes long time to confirm the new crank callers. The second one is the...
limitation of the blacklists’ timeliness. Users have to frequently update the Apps to make sure the blacklist stays up to date. The last one is that the non-smartphone users still suffer the crank calls.

In this paper, we propose and compare two crank calls detection models based on the telephone subscribers’ call data, which can help operators mark crank callers and alert subscribers when they are called. Depending on different ways to organize the features, the two models have distinct frameworks. The first one is a traditional classification model that composes three basic classifiers including decision tree [3], support vector machine (short for SVM) [4], and back propagation neural network (short for BP) [5], in which the ensemble methods Adaboost [6] and voting module are applied to enhance the reliability of the results. The other one is a convolutional neural network (short for CNN) [7] model based on time-sliced data, which contains two sets of convolutional layers and two pooling layers. The experiment on 2 million users’ call data shows that both of the two models are feasible, however, the CNN-based model gets significant higher scores of precision and recall.

2. Model and Method

In this section, we propose and compare two detection models for crank calls, which are based on the call data of telephone subscribers. The two models depend on two different way of organizing the features. Thus, these two models detect the crank calls by applying different frameworks. One is a traditional composite classification model based on statistical features, and the other is a CNN model based on time-slice features that is obtained by dividing the data according to the time-slices.

2.1. Data

Because the proposed crank call detection models are data-driven, we first introduce the basic information of the data.

The data we use in this paper is anonymous call log spanning eight continuous days, which contains 18 basic attributes as show in the table 1, in which, NC represents Not Connected. It is noted that the first six attributes are the statistical features of ringing the bell without making a call.

| No. | Attributes                          | No. | Attributes                          |
|-----|------------------------------------|-----|------------------------------------|
| 1   | Number of active ring g (NC)      | 10  | Number of passive calls            |
| 2   | Duration of active ring (NC)      | 11  | Total passive call duration        |
| 3   | Number of active ring subscribers (NC) | 12 | Number of passive calling subscribers |
| 4   | Number of passive ring (NC)       | 13  | Number of active hanging rings     |
| 5   | Duration of passive ring (NC)     | 14  | Number of passive hanging rings    |
| 6   | Number of passive ring subscribers (NC) | 15 | Number of active hanging calls     |
| 7   | Number of active calls            | 16  | Number of passive hanging calls    |
| 8   | Total active call duration        | 17  | Maximum active call duration       |
| 9   | Number of active calling subscribers | 18 | Maximum passive call duration      |

2.2. Statistical Features Based Classification Model

We first propose the statistical-feature-based detection model. This is a traditional classification model which composites multiple basic classifiers by applying ensemble methods.

Because the number of basic features in original data is small and the difference is insufficient, empirically, we use the basic features to evolve more derivative features before input the data into classifiers. Setting the basic features as the first-order features \( F^{(1)} = \{ f_i^{(1)} \}_{i=1}^{18} \), thirteen composite
features are calculated depending on the first-order features, noted as the second-order features $F^{(2)} = \{ f_i^{(12)} \}_{i=1}^{13}$. For example, average call duration can be calculated as

$$f_1^{(2)} = \frac{f_8^{(1)}}{f_7^{(1)}}$$

(1)

Ratio of active hanging calls can be calculated as

$$f_2^{(2)} = \frac{f_{14}^{(1)}}{f_1^{(1)} + f_7^{(1)}}$$

(2)

Failure rate of active calls can be calculated as

$$f_3^{(2)} = \frac{f_1^{(1)}}{f_1^{(1)} + f_7^{(1)}}$$

(3)

Subscriber diversity can be calculated as

$$f_4^{(2)} = \frac{f_3^{(1)} + f_9^{(1)}}{f_3^{(1)} + f_6^{(1)} + f_9^{(1)} + f_{12}^{(1)}}$$

(4)

And so forth. Furthermore, 39 statistical features are also constructed by combining the first-order features and second-order features, noted as the third-order features $F^{(3)} = \{ f_i^{(3)} \}_{i=1}^{39}$, such as variance of active calls, variance of active-call failure rate, and so forth.

Adaboost [6] is one of the most successful ensemble methods and can combine multiple weak classifier with low prediction accuracy to enhance them as a stronger classifier with higher prediction accuracy. Thus, based on the basic features and derived features, we build a composite classification model to detect the crank calls by applying the Adaboost ensemble method and voting module. As shown in figure 1, three essential classifiers are selected, including Decision Tree [3], SVM [4], BP [5], each of which is enhanced with an Adaboost module. More synthetically, we finally apply the voting module to integrate the results. It is noted that this model is a binary classification model that can identify the crank callers directly from the results.

2.3. Time-slice Features Based Convolutional Neural Network

We next proposed the CNN-based crank call detection model. Instead of using the original data to derive more statistical features, in the CNN-based model, we slice the basic features on a daily base and splice the sliced features into a block. As illustrated in the figure 2, the block represents the expanded matrix based on users’ basic features in multiple time cycles, that is, each row represents a type of feature indicators, each column represents the features in a certain day, thus the corresponding value of $(x, y)$ coordinates is expressed as the value of the $x$-th feature in $y$-th day. As shown in the figure 2, there is a marked difference between the matrix of a normal user (figure 2a) and an abnormal user (figure 2b). For example, the values in 2nd to 5th row of crank call are greater than that in normal users, while, the values in 9th and 10th row of normal users are greater than that in crank calls.

Based on the expanded feature matrix, we set a convolutional neural network with two sets of convolutional layers and two pooling layers, each of which contains 128 filters of size $2 \times 2$. Thus, after four layers of feature abstraction, the original features are expanded into 128×128 abstract features, as illustrated in the figure 3. And then, these abstract features will be calculated by the full connection layer to determine whether the user is an abnormal user or not.
3. Experiments

3.1. Data Description
The data used in this paper is anonymous with 2 million users, in which 0.2 million users are labeled as crank callers. In order to avoid overfitting, we conduct sample balance processing on the training data and sample the positive samples to ensure that the ratio of positive samples (normal users) and negative samples (abnormal users) of the input data is 2:1.
3.2. Evaluation Index

In the experiment, the well-known evaluation indices precision and recall [8] are used to estimate the performance of the two models. Based on the confusion matrix which contains four relationships between benchmarks and predictions as shown in table 2, the precision is calculated as

\[
P = \frac{TN}{FP + TN},
\]

(5)

And the recall is calculated as

\[
R = \frac{TN}{FN + TN},
\]

(6)

where TP is the number of normal users predicted to be normal; FP is the number of abnormal users yet falsely predicted to be normal; FN is the number of normal users yet falsely predicted to be abnormal; TN is the number of abnormal users predicted to be abnormal.

Table 2. The relationships between benchmarks and predictions.

| Predictions | Label: True     | Label: False     |
|-------------|-----------------|-----------------|
| True        | True Positive (TP) | False Positive (FP) |
| False       | False Negative (FN) | True Negative (TN) |

3.3. Results

We compare the two proposed models with 4 benchmark classification algorithms, including Decision Tree [3], SVM [4], BP [5] and KNN [9]. It is noted that the features used in the benchmark algorithms are statistical features which are same as the proposed composite classification model. In order to ensure the validity of the experiment, 10 cross-validation method is applied to evaluate the average precision and recall.

As shown in the table 3, the simple classification algorithms get unsatisfactory results, although the precision of KNN is higher than 90% that is close to the composite classification, its recall is far too low, while both of the two proposed models get qualified results, the precisions are higher than 90%. Furthermore, the CNN-based model performs overall better than the composite classification, the precision of CNN-based model is close to 100%, and the recall of the CNN-based model is much higher than composite classification by nearly 10%.

Table 3. The precision and recall of the two proposed model.

|                  | Precision | Recall  |
|------------------|-----------|---------|
| Decision Tree    | 78.90%    | 70.26%  |
| SVM              | 72.21%    | 59.88%  |
| BP               | 73.41%    | 66.97%  |
| KNN              | 91.11%    | 42.98%  |
| Composite Classification | 91.20%    | 71.77%  |
| CNN-based Model  | 99.32%    | 82.09%  |

4. Discussion

Based on the call data, we propose two types of model to detect the crank calls. The traditional composite one, without surprise, gets satisfactory precision, while the CNN-based one gets remarkable
performance, the precision is closed to 100%. Essentially, the excellence of the CNN-based model lies in the way it organizes the data. The spliced time-sliced data can bring much more diverse abstract features, which is achieved by using convolution operations and filters.

However, the recalls of these two models are not good enough, there are still many aspects in our models that can be optimized, especially the CNN-based model. First of all, the optimization of features can be considered. The improvement of features is extremely important, which needs to be improved in terms of the number of basic features, the number of manual statistical features, the time span, and the time granularity. Secondly, it can be considered that the model parameters should not adapt to the situation. Therefore, parameters in the CNN-based model need to be fine tuned, including the number of convolution and pooling layers, the size and number of filters, etc.

5. Acknowledgments
This work is partially supported by the National Natural Science Foundation of China under Grant No. 61703074, and by the Fundamental Research Funds for the Central Universities under Grant No. ZYGX2016J196, and by Natural Science Foundation of Jiangsu Province under Grant No. BK20180209.

6. References
[1] 2016  Analysis report of China’s crank call situation in 2016 (Shanghai: CooTek Report)
[2] 2019  Market analysis of the development status and governance trend of China’s crank calls in 2019 (Guangzhou: iiMedia Research Report)
[3] Quinlan J R 1994  Mach. Learn. 16 235-240
[4] Vapnik V 1995  The nature of statistical learning theory (New York: Springer-Verlag)
[5] Rumelhart D E and McCelland J L 1986  Parallel Distributed Processing (Cambridge: MIT Press)
[6] Freund Y and Schapire R E 1997  Journal of Comp. Syst. Sci. 55 119-139
[7] LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W and Jackel L D 1989  Neural Comp. 1 541-551
[8] Antonio M, Petraq J P, Panos M. P 2006 Validation  Data Mining in Agriculture (New York: Springer) chapter 8 pp 161-172
[9] Antonio M, Petraq J P, Panos M. P 2006 Validation  Data Mining in Agriculture (New York: Springer) chapter 4 pp 83-106