International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n3.34
Journal homepage: https://wepub.org/index.php/IJCSIT/index

# Machine Learning in the Application of Fraudulent SMS Filtering

Yihe Huang *

Department of Hubei University, Hubei, China
*Corresponding Author: 2479955364@qq.com

## ABSTRACT

Electronic fraud SMS is a common network security threat, which brings huge economic and privacy risks to users. To effectively address this problem, this paper studies the electronic scam SMS filtering method based on machine learning. By analyzing and processing a large number of SMS datasets, we propose a classification model that comprehensively considers both textual and behavioral features to identify and filter e-scam text messages. The experimental results show that our method achieves significant improvements in accuracy and efficiency and provides better security for users.

## KEYWORDS

Electronic fraud SMS filtering; Feature selection; HHO-KNN algorithm; Intuitive fuzzy set

## 1. INTRODUCTION

With the rapid development of the electronic information industry and the increase of e-commerce business, network users receive a large number of text messages every day, which is mixed with a large number of pornographic and violent advertising. Some criminals have targeted some pretty cheap people and some can not distinguish the elderly, to them to telecom fraud to obtain illegal money. For example, the movie "All the eggs" released this summer vacation tells the story of a fraud ring in northern Myanmar. One of them is a well-off family ruined because their children were cheated. Such examples are not a few, telecom fraud has led to a lot of family breakdown, but there are still a lot of people can not distinguish, so I want to use some programs to filter out the telecom fraud calls can reduce the loss of personal property.

In order to solve the problem of fraudulent SMS classification, many scholars and researchers have conducted a lot of research work. Among them, the classification model based on machine learning algorithm is one of the most common and effective methods. However, traditional machine learning algorithms are faced with challenges such as difficulty in feature selection and lack of model generalization ability in handling the problem of fraudulent SMS classification.

In order to overcome the shortcomings of traditional machine learning algorithms, this paper proposes a method to construct a fraudulent SMS classification model based on the HHO-KNN algorithm. This method combines the HHO-KNN algorithm to improve the performance of the fraudulent SMS classification model by optimizing feature selection and classification accuracy. Specifically, this paper first uses HHO algorithm for feature selection, transform the original feature space into subspace with better discrimination; then, using KNN algorithm to realize accurate identification of fraudulent SMS.

For the construction of fraudulent SMS filtering model, this paper proposes a method to construct fraudulent SMS filter model based on three-branch decision ideas and intuitive fuzzy set. This method starts from three perspectives, including decision ideas, intuitive fuzzy set and multi-feature fusion, and constructs a fraud SMS filter model that comprehensively considers different factors. By introducing multiple attributes and weights, as well as the design of fuzzy membership functions, the model can filter out fraudulent text messages more accurately, and improve the user's sense of security and use experience.

The electronic fraud SMS filtering problem can be converted into two sub-problems: how to select a feature subset that can best represent the original data set; the second is how to construct a reasonable and accurate SMS classifier.

For the first subproblem, the HHO-KNN optimization algorithm.

For the second question, this question with two decision ideas and intuition fuzzy set, study a classification model for telecom fraud SMS filtering, and then for intuition fuzzy set, study a classification framework for telecom fraud SMS filtering, and combined with the similarity measure and two decision ideas, study a cost sensitive telecom fraud SMS filter.

## 2. CONSTRUCTION OF FRAUD SMS CLASSIFICATION MODEL

In recent years, scholars at home and abroad have conducted extensive research on the classification of fraudulent SMS. Among them, the classification methods based on machine learning algorithms are widely used. For example, algorithms such as SVM and Naive Bayes are used to build fraudulent SMS classification models. Although these algorithms have achieved some results to some extent, there are still some problems in practical application. Therefore, further research and improvement of existing algorithms are needed to improve the accuracy and efficiency of fraudulent SMS classification.

We used the HHO-KNN algorithm for the classification and feature selection. Text features include word frequency, word vector, and grammatical features, while behavioral features include sending frequency, sending time, and sender identity.

The algorithm process is divided into pre-processing stage, text representation and feature selection:

(1) The preprocessing stage text cleaning is used to handle the noise and unnecessary characters contained in the data, as well as letter case conversion and deletion of numbers, punctuation, spaces, links and stop words. Data is converted into a digital format by the bag of word model (BoW) and the word frequency-reverse text frequency index TFIDF.

(2) Text presentation stage at this stage, we use VCS for text representation, giving the following two messages:

a. He likes watching movies very much, and she also likes watching movies very much.

b. He also likes to watch sports games.

The above two sentences appear a total of 14 different words {he, she, very, very, also, also, happy, joy, joy, see, electricity, electricity, shadow, body, education, yu, comparison, competition,}, based on the above two sentences, using the index of the thesaurus to indicate the vector of length 14:

a. {1, 1, 1, 0, 1, 2, 2, 2, 2, 2, 0, 0, 0, 0}

b. {1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1}

The number indexed in each vector indicates how many times the word in that aurhas appears. For example, "she" appears once in the first sentence so that the second position element of the first vector

is "1", while the second sentence does not appear "she" so that the second position element of the second vector is "0".

An SMS is expressed as a document and each in the SMS does not repeat the word as a feature item. Here, we use the TF-IDF method to calculate the weights:

W (a, b) = TF (a, b) * log (N/na+1)

Where w (a, b) represents the weight of a word in the text b, TF (a, b) represents the frequency of the word a in the text b, N is the total number of trained texts, naRepresents the number of texts where the word a appears in the training text.

(3) Feature selection (FS) uses the HHO optimization algorithm as a method to select a subset of features. Feature selection acts as the storage form via a vector consisting of t and f, t indicates that the attribute is selected and f indicates unselected. The number of features included in the solution is the same as the number of properties of the dataset.

The steps of the HHO-KNN optimization algorithm are as follows:

Step 1: Read the data;

Step 2: Preprocess the data;

Step 3: Initialize the hawks;

Step 4: Determine whether the iteration number reaches the T;

Step 5: execute the HHO algorithm;

Step 6: calculate the fitness of all solutions using the KNN classifier;

Step 7: Find the highest fitness score as the local optimal solution;

Step 8: Repeat until the number of iterations reaches T;

Step 9: Record the local optimal solution as the global optimal solution, and output it.

In the HHO-KNN optimization algorithm, after each iteration, the fitness of the attribute set is recalculated, the local best fitness is selected, and then the iteration and counting are repeated, and the fitness is calculated until the number of iterations is reached or the required fitness is obtained, and the final attribute set is used as the best special collection. In each iteration of HHO, the value of the feature is constantly modified and the value of fitness is increased by KNN until the optimal solution is obtained

According to the literature "Spam Detection Research based on HHO-KNN Optimization Algorithm":

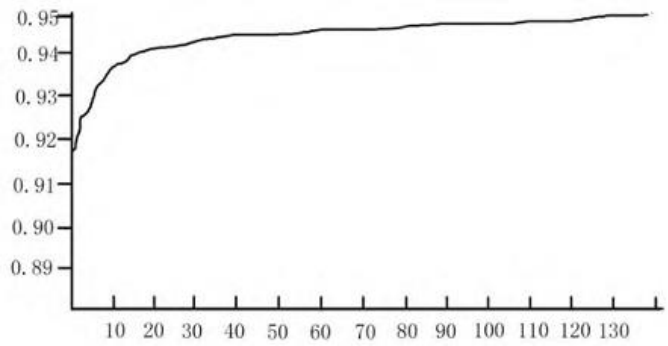


**Figure 1.** Relationship between accuracy and number of iterations of the HHO-KNN algorithm

Note: The vertical axis represents accuracy, and the horizontal axis represents Number of iterations

It can be seen that the accuracy of the HHO-KNN algorithm increases with the number of iterations, and the accuracy increases gently and tends to 0.95.

**Table 1.** Fitness of several algorithms on the datasets

| metric | Dragonfly algorithm | Balanced optimization algorithm | Seagull optimization algorithm | The marine predator algorithm | HHO-KNN algorithm |
|---|---|---|---|---|---|
| The best value | 45.3146 | 45.1312 | 44.3835 | 44.6403 | 44.3583 |
| average value | 49.2286 | 49.2358 | 50.4390 | 48.7403 | 44.3531 |
| The worst value | 66.2349 | 62.6545 | 54.4922 | 56.4403 | 44.3554 |
| standard value | 9.0564E-033 | 1.1414E+014 | 4.1549E-022 | 1.1357E-101 | 1.0857E-101 |
| It can be seen that the best and worst values of HHO-KNN are lower than other algorithms, and the fluctuation between the best and worst values is small | | | | | |

The clustering accuracy of each algorithm is improving after the increase of iterations, but at the end of the iterations, the clustering accuracy is higher than that of other algorithms, and the iteration process converges

Conclusion: The KNN-HHO algorithm proposed in this paper has a better effect in the feature selection of telecom fraud SMS than in the other algorithms. The algorithm can help us to select the most representative and distinguishing features from the large feature set, and improve the classification accuracy of telecom fraud short messages. In the future, we will further improve the algorithm and conduct experiments to verify its feasibility and effectiveness on a larger dataset.

Outlook: This paper constructs a fraud SMS classification model based on HHO-KNN algorithm, and verifies the accuracy and effect of the model through experiments. However, in practical applications, there are still some problems to solve further. Future research can be conducted from the following aspects: further optimizing the parameter setting of the HHO-KNN algorithm, further expanding the scale of the training set, and applying the model to the practical SMS platform. It is hoped that through continuous research and improvement, the accuracy and efficiency of the fraudulent SMS classification model can be improved, and a better SMS security guarantee can be provided for users.

## 3. CONSTRUCTION OF THE FILTER

The construction of filters is essentially a classification problem, and we can construct a function f that maps continuous type variables into discrete type variables. We will use VSMS for text representation with a feature vector, namely X= (X1, X2, X3....., Xm) (m means that there are m features in the SMS message) and then classify them with the function f.

### 3.1. Intuitive Fuzzy Set

Let X= {x1, x2, ...., Xn} is a non-empty set, called F= {<x, uA(x), vA(X)> | x X} is the fuzzy set, where uA, vAIs the membership function and non-membership of the fuzzy set F, uA: X→[0, 1] x∈ X→uA[0, 1], vA: X→[0, 1] x∈X→vA[0, 1], and 0 <=uA+vA<=1, x∈X.

### 3.2. Two Decision-Making Ideas

Let U= {x1, x2, ....., Xn} is a set of finites, non-empty entities (objects) or decision schemes, C is a set of finite or non-empty entities (objects) or a finite set of conditions that may contain indicators, targets or constraints. Based on the conditional set C, the main task of the two-branch decision is to divide the entity set U into two disjoint domains through the mapping function f, respectively classified as POS and NEG, which are called positive and negative domains respectively (as shown in the figure below). For domains of two-branch decision making, acceptance and rejection will be used separately.

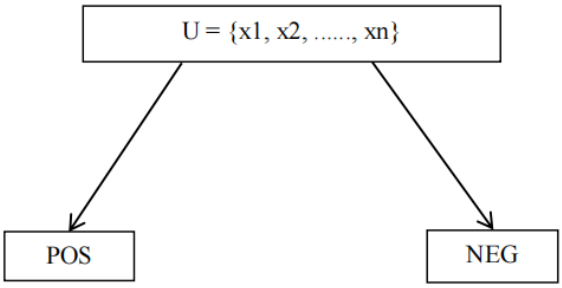


**Figure 2.** Flow chart of two decision-making ideas

### 3.3. Principle of Filter

The electronic fraud SMS filtering problem is essentially a classification problem. There are two forms of electronic fraud SMS: one is harmful or meaningless to users, which may cause great harm to users, so the SMS needs to be removed; the other is the normal SMS harmless to users, that is, legal SMS, which does no great harm to users, so the SMS needs to be retained. Therefore, to establish an electronic fraud SMS filter is to build an electronic SMS classification model. The classical electronic fraud SMS filtering model is a two-way classification model, and the general method makes a clear division between electronic fraud SMS and legal SMS, that is, "either / or".

We take the data from the first part after preprocessing and text representation, nabIt represents the number of times the word a appears in text b and then expressed in VSM. Then calculate it: first standardize and then calculate the membership and non-membership.

We chose the mean and standard deviation for normalization, let nab$\bar{n}$The mean value is, and the standard deviation is SabThen, the standardized formula is: xab=(nab-$\bar{n}$)/Sab

Then we calculate the membership and non-membership by weighting: weight $r_a=|(1-S_{ab})/(1+S_{ab})|$, $u_{ab}=r_a*1/(1+e^{-x_{ab})}$, $v_{ab}=r_a*1/(1+e^{x_{ab})}$.

We just build a set of feature sets X = {($u_{a1}$, $v_{a1}$), ($u_{a2}$, $v_{a2}$), ...., ($u_{ab}$, $v_{ab}$).}

It contains two categories C= {C1, C2}, and we calculate the mean membership and mean non-membership of $t_i$ features belonging to Cn in each SMS separately:

$$\overline{u_{ab}} = \frac{\sum(u_{abX_n(b)})}{\sum X_n(b)}$$

$$\overline{v_{ab}} = \frac{\sum(v_{abX_n(b)})}{\sum X_n(b)}$$

$$\mathrm{Xn(b)} = \begin{cases} 1, \mathrm{b} \in \mathrm{Cn} \\ 0, \mathrm{Other} \end{cases}$$

We will set the learning style as pn = {($\overline{u_{ab}}$, $\overline{v_{ab}}$)}

For a new SMS, we can also calculate its average membership and average non-membership and then classify it.

Entity set B= {b1, b2, ....., The bn} represents the set of email messages to be classified, exists the state set S= {T, F} has two states, exists the classification pattern corresponding to each state P= {pT, pF}. Its decision function is defined as follows:

f (b)=V (pT, b)-V (pF, b)

V represents a similarity measure.

Let's set a = ASC (| f (b)|)$_{ij}$

Then a represents the ascending order of the set and the number of i. j

Then the positive domain POS= {b | f (b)> a};

Negative domain NEG = {b | f (b) <=a};

Where the positive domain indicates the receptive field, and the negative domain indicates the rejected domain.

For the algorithm implementation, I found the following algorithm by searching the literature:

1: / / ① To divide the training set and the test set

2: divide and obtain the training set Dtrain, test set Dtest, And separately containing ntrainAnd ntestSeal the mail

3: / / ② Through the training set DtrainObtain classification patterns,

4: FOR k=1,2 DO

5: FOR i=1,2...M DO

6: Calculate the average membership $\overline{u_{ab}} = \frac{\sum(u_{ab}X_n(b))}{\sum X_n(b)}$ and average non-membership $\overline{v_{ab}} = \frac{\sum(v_{ab}X_n(b))}{\sum X_n(b)}$ in e-mail

7: Calculate the average hesitation of the features in the email

8: END FOR

9: END FOR

10: Get the classification pattern, pn={$(\overline{u_{ab}}, \overline{v_{ab}})$}

11: / / ③ Calculate the pn take part in Dtest The similarity measure of the email in

12: FOR t=1, 2, ...., ntest DO

13: Select a certain similarity measure method to calculate the similarity measure V of the first email and classes C1 and C2 (pT, B) and V (pF, b)

14: Calculate the difference of the similarity measureΔt

15: END FOR

Δt16: rank all in ascending order to obtain position ntestpThe number of ntestpAs a classification threshold for the delayed mail classΔ

17: / / ④ by mode pn classify

18: Set up an empty set Ctest Used to store the test set was pn Classification of labels

19: FOR t=1, 2, ...., ntest DO

20: IF > AND V (pΔtΔT, b)>V (pF, b) THEN

21: Ctest Add C1 to the middle

22: ELIF log (N/na+1)Δ> AND V (pT, b) <V (pF, b) THEN

23: Ctest Add C2

24: END IF

25: END FOR

26: / / ⑤ Computational model performance

27: initialize nss, nLL, nLS And nSL Is 0, which is used to count the number of spam correctly classified, legitimate mail correctly classified, legitimate email incorrectly classified and spam misclassified, respectively

30: FOR t=1, 2, ...., ntest DO

31: IF Ctest [t]=labeltest [t]=1 THEN

32: nss+=1

33: ELIF Ctest [t]=labeltest [t]=2 THEN

34: nLL+=1

35: ELIF Ctest [t]=1 AND labeltest [t]=2 THEN

36: nLS+=1

37: ELES

38: nSL+=1

39: END IF

40: END FOR

41: Calculate classification accuracy, comprehensive evaluation index, weighted classification, weighted comprehensive evaluation index and total cost rate

42: Output the model classification performance index

The results are as follows:

**Table 2.** Experimental results of IFSs-3WD under different parameter designs

| performance index SM | | $V^2_{LS}$ | $V_X$ | $V_Y$ | $V_S$ | $V_{SM}$ |
|---|---|---|---|---|---|---|
| avg_Acc | | 89.31 | 89.27 | 90.29 | 90.68 | 90.44 |
| avg_F1 | | 90.25 | 90.19 | 91.19 | 91.46 | 91.25 |
| W=1 | P=0.1 | 92.76 | 93.06 | 93.67 | 94.39 | 93.47 |
| | WAcc | 91.82 | 92.22 | 92.63 | 93.71 | 92.97 |
| | WF1 TCR | 6.27 | 6.50 | 7.02 | 8.02 | 6.92 |
| | P=0.15 | 94.19 | 94.19 | 94.84 | 94.95 | 94.62 |
| | WAcc | 93.34 | 93.62 | 94.12 | 94.51 | 94.09 |
| | WF1 TCR | 7.74 | 7.80 | 8.65 | 9.02 | 8.48 |
| | P=0.2 | 95.40 | 95.06 | 96.32 | 96.21 | 96.43 |
| | WAcc | 94.48 | 94.27 | 95.69 | 95.89 | 95.98 |
| | WF1 TCR | 9.57 | 9.02 | 12.13 | 12.09 | 12.74 |
| | P=0.5 | 97.22 | 97.04 | 98.89 | 99.44 | 99.44 |
| | WAcc | 93.60 | 92.61 | 98.98 | 99.48 | 99.42 |
| | WF1 TCR | 8.53 | 8.44 | 47.67 | 93.67 | 87.83 |
| W=3 | P=0.1 | 94.37 | 94.61 | 94.59 | 95.12 | 94.95 |
| | WAcc | 87.68 | 87.87 | 87.21 | 89.40 | 89.04 |
| | WF1 TCR | 3.86 | 4.05 | 3.92 | 4.52 | 4.29 |
| | P=0.15 | 96.20 | 95.26 | 96.17 | 95.91 | 95.98 |
| | WAcc | 90.95 | 89.90 | 91.36 | 91.38 | 91.60 |
| | WF1 TCR | 5.68 | 4.54 | 5.57 | 5.45 | 5.46 |
| | P=0.2 | 96.57 | 96.58 | 96.65 | 97.01 | 97.22 |
| | WAcc | 91.99 | 92.09 | 92.50 | 93.52 | 93.66 |
| | WF1 TCR | 6.08 | 6.23 | 6.31 | 7.42 | 7.89 |
| | P=0.5 | 98.62 | 98.70 | 99.15 | 99.72 | 99.53 |
| | WAcc | 90.61 | 93.53 | 98.48 | 99.47 | 99.12 |
| | WF1 TCR | 6.99 | 7.68 | 32.59 | 95.52 | 51.90 |
| W=9 | P=0.1 | 94.88 | 94.97 | 95.40 | 95.76 | 96.06 |
| | WAcc | 78.91 | 77.77 | 78.15 | 79.89 | 81.51 |
| | WF1 TCR | 1.66 | 1.71 | 1.80 | 2.02 | 2.13 |
| | P=0.15 | 96.49 | 96.47 | 96.48 | 96.58 | 96.51 |
| | WAcc | 81.55 | 84.19 | 82.55 | 84.75 | 84.32 |
| | WF1 TCR | 2.45 | 2.41 | 2.37 | 2.55 | 2.49 |
| | P=0.2 | 97.99 | 97.55 | 97.51 | 98.21 | 97.63 |
| | WAcc | 88.33 | 87.34 | 87.79 | 91.05 | 89.27 |
| | WF1 TCR | 4.07 | 3.37 | 3.33 | 4.91 | 3.70 |
| | P=0.5 | 99.57 | 99.55 | 99.48 | 99.61 | 99.60 |
| | WAcc | 92.59 | 92.46 | 98.08 | 98.31 | 97.99 |
| | WF1 TCR | 8.71 | 7.73 | 22.50 | 28.84 | 24.64 |

It can be seen that the algorithm has good classification performance, which can effectively improve the performance of the classification model, can provide users with a suitable electronic SMS classification reference, and save time cost for users.

## 3.4. Summary and Outlook

This chapter proposes a text classification framework based on intuitive fuzzy sets, which shows good performance in classification tasks. The algorithm classifies email based on the similarity measure of the intuitive fuzzy set and combines the idea of two-branch decision to classify and filter SMS.

There are still many shortcomings to be improved: two decision ideas without three decision ideas can be classified and filtered, and rely on the training set and test set, which can be improved in future algorithm improvements.

## REFERENCES


[1] Research on spam detection based on HHO-KNN optimization algorithm_Chen Liang.pdf

[2] Aljarah, I., Faris, H., & Mirjalili, S. (2019). Hybrid krill herd algorithm with k-nearest neighbors for data classification. Applied Soft Computing, 74, 166-175.

[3] Aljarah, I., Faris, H., & Mirjalili, S. (2018). Optimizing the training parameters of support vector machines using hybrid harmony search algorithm. Neural Computing and Applications, 29(11), 1073-1086.

[4] Aljarah, I., Faris, H., & Mirjalili, S. (2017). HHO-KNN: A hybrid harmony search algorithm with k-nearest neighbors for data classification. Expert Systems with Applications, 77, 236-250.

[5] Li, X., Zhang, Y., & Zhang, J. (2019). Feature selection for spam detection using a hybrid KNN-HHO algorithm. Journal of Computational Science, 31, 140-148.

[6] Wang, X., & Li, X. (2018). A novel hybrid KNN-HHO algorithm for feature selection. PloS one, 13(5), e0197396.

[7] Liu Cheng, Chen Xi, Wu Wenbo. Research on spam SMS analysis and management technology [J]. Guangdong Communication Technology, 2022, 42 (06): 16-20.

[8] Li Shu. Escape attack and its application in Chinese spam message filtering [D]. Nanjing University of Posts and Telecommunications, 2018.

[9] Zhang Yan, Liu Yin, Liu Shengping, etc. Research on intelligent governance system for spam SMS [J]. Telecommunications Network Technology, 2017, No.282(12):67-71.

[10] Xia, Tian, and Xuemin Chen."A Discrete Hidden Markov Model for SMS Spam Detection." Applied Sciences (2020).