Research on Feature Selection Algorithm of Energy Curve

Xiaohong Fan[1], Ye Huang[1], Xue Wang[2], Ziran Nie[1], Zhenyang Yu[1], Xuhui Cheng[1], and Xiaoyi Duan[1(✉)]

[1] Beijing Electronic Science and Technology Institute, Fengtai District, Beijing, China
xiaoyi_duan@sina.com

[2] Beijing Electromechanical Engineering Research Institute, Fengtai District, Beijing, China

Abstract. Energy analysis attack is a side channel attack, which collects and analyzes the power leakage information in the operation process of cryptographic chip, and then recovers the correct key. In the process of energy analysis attack, the collected power leakage information has many feature dimensions and a large amount of data. Putting all the features into the algorithm will bring dimension disaster. Therefore, choosing the characteristic points of the energy curve is of great significance for the success of the attack. Firstly, three kinds of feature selection methods are studied in this paper. Secondly, three energy curve feature selection algorithms are implemented: dynamic feature selection algorithm based on mutual information, feature selection algorithm based on decision tree and feature selection algorithm based on recursive feature elimination. Finally, the three feature selection results are tested and evaluated by machine learning, which shows that the subsets generated by the three algorithms have good performance and can be used for energy analysis attacks. Among the three methods, the feature selection algorithm based on decision tree has a short-time and the selected feature subset is the best.

Keywords: Energy analysis attack · Feature selection · Mutual information · Decision tree · Recursive feature elimination

1 Introduction

Energy analysis attack is an effective attack method against cryptographic chip. In recent years, machine learning has developed rapidly and has been applied to side channel attacks, which has achieved excellent results. In the energy analysis attack, the collected energy curve data has high dimension, large magnitude, and contains a large amount of irrelevant and redundant information. Thus, the machine learning algorithm needs to deal with a large amount of data, which increases the difficulty of energy analysis attack and reduces the attack efficiency. Therefore, in the process of energy analysis attack, it is necessary to select high-value feature points from a large number of data sets, that is, feature selection of energy curve, which is a key step in energy analysis attack.

This paper is supported by "the Fundamental Research Funds for the Central Universities" (Grant Number:328202207, 328202247, 3282023054).

S. Goel and P. R. Nunes de Souza (Eds.): ICDF2C 2023, LNICST 570, pp. 293–307, 2024.
https://doi.org/10.1007/978-3-031-56580-9_18

Feature selection is to select the best feature subset from the original feature set. It is one of the key technologies in the field of machine learning and data mining, and its stability and efficiency are also the current research hotspots. So far, many scholars have defined feature selection from different perspectives: Kira et al. define that feature selection is to find the minimum feature subset necessary and sufficient to identify the target in an ideal situation [1]; John et al. defined feature selection from the perspective of improving prediction accuracy as a process that can increase classification accuracy or reduce feature dimension without reducing classification accuracy [2]; Koller et al. defined feature selection from the perspective of distribution as selecting a feature subset as small as possible under the condition that the distribution of the result class is similar to that of the original data class [3]. The above definitions have different starting points and focus, but their goal is to find a minimum feature subset that can effectively identify the target. In 1996, Kocher et al. proposed a feature selection algorithm based on mutual information [4]; In 2003, the method based on mean difference was introduced [5]; In 2006, Gierlichs et al. proposed the method based on paired t-difference sum of squares and difference sum of squares [6]; In 2019, Archambeau et al. proposed the method of PCA (Principal Component Analysis) [7], and in the same year, aiming at the imbalance of data categories, Ireneusz et al. used stack technology to improve the generalization ability of machine learning in feature selection [8]; In 2020, Khosla et al. introduced TPDS (Topologically Preserved Distance Scaling) to strengthen feature selection, so as to reproduce distance information in a higher dimension. Compared with typical distance preservation methods, TPDS can provide better visualization and better classification of data points based on narrowing feature points [9].

In the current research on feature selection, in addition to pursuing efficient learning performance, scholars also pay attention to the stability of feature selection, that is, insensitive to the changes of training samples. At present, there are two methods to improve the stability of feature selection. One is to give different weights to the training samples, so that the result of feature selection can get a better balance between variance and deviation. The other is to use the idea of ensemble learning and use ensemble feature selection to effectively improve the stability of feature selection. Although there are many feature selection methods, there are still many deficiencies in solving practical problems. It is necessary to select the required metrics and classifiers according to the specific environment.

This paper mainly studies three energy curve feature selection methods based on filtering, wrapper and embedded mode, and realizes the effective feature point selection for energy analysis attacks. Through the comparison of these three different feature selection algorithms, a relatively efficient and accurate feature selection algorithm is obtained.

2 Principle of Feature Selection Algorithm

Feature selection refers to the process of selecting the most effective features from a group of features to reduce the dimension of feature space. It is a data preprocessing technology in the field of machine learning. Feature selection can screen out the irrelevant and redundant features in the data set, so as to reduce data storage requirements,

reduce machine learning model training time, and improve model prediction ability. The basic framework of feature selection is shown in Fig. 1, which mainly includes four parts: generating subsets, evaluation criteria, stop conditions and result verification [10]. According to literature [1–3], generating subsets and evaluation criteria are the core parts, which complete the main work of feature subsets screening. According to the evaluation criteria, feature selection algorithms can be divided into three types: filtering, wrapper and embedded.

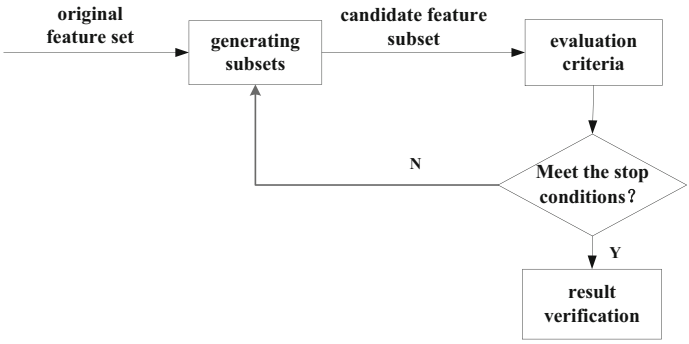

Fig. 1. Basic framework of feature selection

2.1 Filtering Mode

The filtering mode scores the data set according to the divergence or correlation, and sets the number of the threshold or the threshold to be selected, so as to select the features that meet the conditions. Then, trains the classifier, and the feature selection process is independent of the subsequent classifiers [11]. The principle of filtering mode feature selection is shown in Fig. 2.This mode can be used as a good feature preselector, which has high computational efficiency, can quickly eliminate a large number of non-critical noise features, and narrow the search range of the optimized feature subset. However, there is also a major problem in this mode, that is, it is unable to select a small-scale optimal feature subset, especially when the feature and classifier have a large association. The filtering mode has a variety of evaluation criteria, including distance measurement, information measurement, and correlation measurement and so on. Different evaluation criteria lead to different feature subsets.

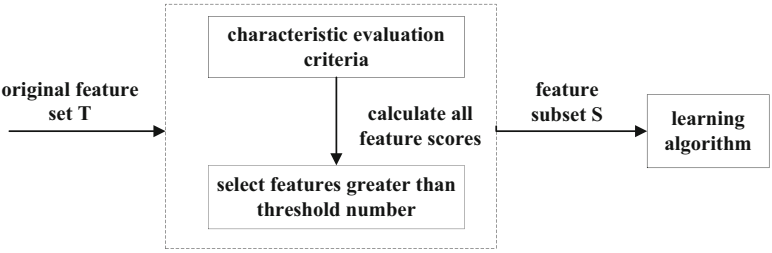

Fig. 2. Schematic diagram of filtering mode feature selection

2.2 Wrapper Mode

The wrapper mode integrates the feature selection with the classifier training, which are completed in the same optimization process. That is, feature selection is automatically performed during the classifier training process. The principle of wrapper mode feature selection is shown in Fig. 3.The performance of this mode is closely related to the classifier it uses. In the process of filtering features, it directly uses the selected features to train the classifier, and evaluates the selected features according to the performance of the classifier on the verification set. Because each candidate subset needs to be retrained, this mode runs longer than the filtering mode, which is not conducive to high-dimensional and large sample data. However, the selected optimized feature subset is relatively small, which is more suitable for a specific learning algorithm [12].

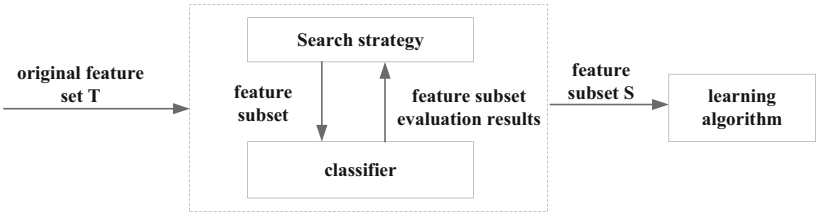

Fig. 3. Schematic diagram of wrapper mode feature selection

2.3 Embedded Mode

Embedded mode is to use some machine learning algorithms and models for training to obtain the weight coefficients of each feature, and select features from large to small according to the coefficients. The principle of embedded mode feature selection is shown in Fig. 4. In this mode, the feature selection algorithm is embedded in the learning algorithm, and the feature subset can be obtained through the training of the classification algorithm. Embedded mode is similar to filtering mode, but it determines the quality of features through training. It can solve the problem of high redundancy of filtering mode results and long running time of wrapper mode, which can be regarded as a compromise of the other two modes. The embedded mode has high efficiency, but its shortcomings

are also obvious. The feature subset selected by this mode has excellent performance, but only for itself, and is prone to over fitting.

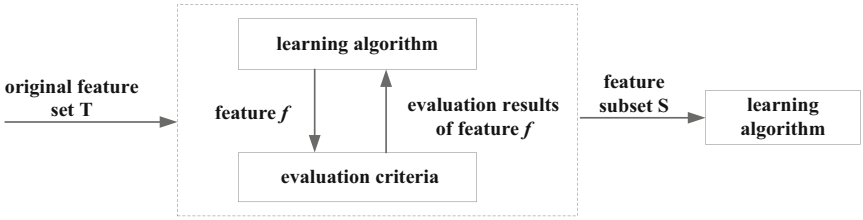

Fig. 4. Schematic diagram of embedded mode feature selection

Generally speaking, filtering mode takes a short time, but subset selection depends on specific evaluation criteria. Wrapped mode subset has good performance, but it is highly dependent on the learning algorithm and is not suitable for high-dimensional data. Embedded mode has high efficiency and good subset. It can process high-dimensional data, but it is prone to over fitting. There are many kinds of feature selection methods, and different methods have different defects. Therefore, it is necessary to select the appropriate scheme according to the actual situation of the research object.

3 Implementation of Feature Selection Algorithm

The data set used in this paper is from the international academic competition of differential energy analysis. The masked AES-256 on ATMega163 chip is taken as the analysis object. A total of 10000 energy curves are collected, and each curve carries 600 characteristic points.

The energy analysis attack on the masked AES-256 encryption algorithm is to collect the energy leaked during the operation of the algorithm, analyze and obtain the Hamming weight of the first S-box output in step SubBytes of the algorithm, and then get the first byte of the key used. Therefore, the purpose of data feature selection is to select the feature points in each energy curve that are highly related to the Hamming Weight of the first S-box output from the original data set.

3.1 Dynamic Feature Selection Algorithm Based on Mutual Information

Assuming that X and Y are two continuous random variables, $p(xy)$ is the joint probability density function, $p(x)$ and $p(y)$ are the marginal density functions, the mutual information $MI(X;Y)$ between X and Y can be obtained as Formula (1).

$$MI(X;Y) = \int\int p(xy)\log\frac{p(xy)}{p(x)p(y)}dxdy \tag{1}$$

MI (Mutual Information) can also be regarded as the difference between entropy and conditional entropy, that is, the reduction of uncertainty of Y under the given condition of X, as shown in Formula (2).

$$MI(X;Y) = H(Y) - H(Y|X) \tag{2}$$

If X can determine Y, then $H(Y|X) = 0, MI(X;Y) = H(Y)$. If X and Y are not related to each other, then $H(Y|X) = H(Y), MI(X;Y) = 0$.

Similarly, conditional mutual information can be understood as the mutual information between variables X and Y under the given variable Z, as shown in Formula (3).

$$MI(X;Y|Z) = H(X|Z) + H(Y|Z) - H(XY|Z) \tag{3}$$

MI can measure the relationship between variables, which is invariant under spatial transformation. MI can be used for feature selection, through which the features that are highly related to classes but have low redundancy with other features can be screen out [8]. This paper first uses MI to select 10000*60 feature subsets from the original data set; then uses dynamic feature selection algorithm of MRIDFS to select 10000*10 feature points from the redundant data set; finally a 10000*70 data set is formed. The algorithm flow is shown in Fig. 5.

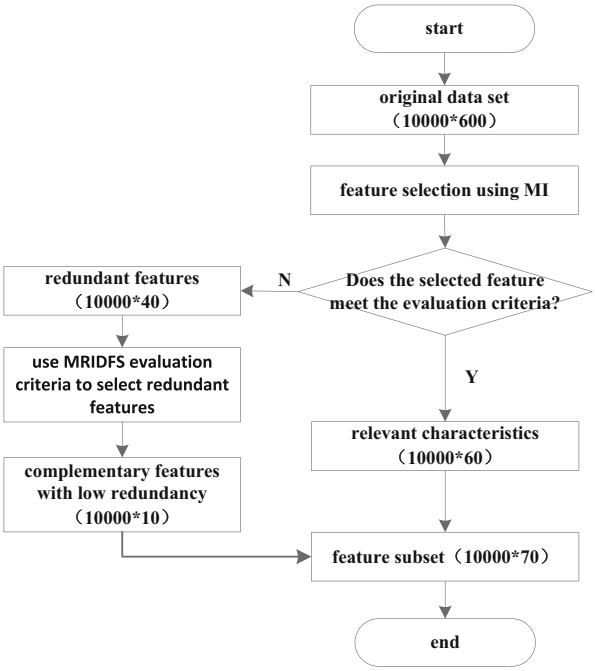

Fig. 5. The flow of feature selection algorithm based on MI

On the basis of DCSF algorithm, MRIDFS algorithm introduces the concept of the feature-dependent redundancy ratio to dynamically adjust the importance of the redundancy out of the class [9]. The evaluation function of MRIDFS algorithm is Formula (4).

$$\mathrm{J}(X_i) = \sum_{X_s} I(X_i;Y) + I(X_i;Y|X_s) + I(X_s;Y|X_i) - \sum_{X_s \in S} \frac{I(X_i;X_S)}{I(X_i;Y)} I(X_i;X_s|Y) \tag{4}$$

where X_i is the candidate feature, X_s is the selected features and Y is the class label. Ratio $\frac{I(X_i;X_S)}{I(X_i;Y)}$ in Formula (4) is defined as the feature-dependent redundancy ratio, which refers to the redundancy attached to per unit related information. The smaller the ratio, the lower the actual redundancy.

3.2 Feature Selection Algorithm Based on Decision Tree

Decision tree is a very common classification method. Each internal node represents an attribute test, each branch represents a test output, and each leaf node represents a category. Theoretically, the tree model can be used to divide the feature space infinitely. At the same time, regularization item and pruning strategies that reflect the complexity of the tree model can be added to prevent over fitting. Therefore, the variance and deviation can be weighed by parameter adjustment to obtain better accuracy. In decision tree, the construction process of the tree is the process of feature selection. Decision tree is widely used in both classification and regression problems. However, a single tree model is not commonly used in data analysis. In the current data algorithm field, GBDT, XGBoost and LightGBM are three popular composite models based on tree.

GBDT, also known as gradient lifting tree, is an additive model based on Boosting, and its base model is CART. It uses decision tree iterative training to get the optimal model, which has good training effect and is not easy to over fit. However, GBDT needs to traverse the entire training data many times in each iteration step. For massive data, embarrassingly, if the entire training data is written into the memory, the size of the training data will be limited; if the training data is not written to the memory, it will take a lot of time to read and write the training data repeatedly. XGBoost is a method based on pre-ordering, which is improved on the basis of GBDT. It is excellent in algorithm optimization and system implementation, and performs better than the traditional GBDT in speed and accuracy. However, it faces the similar problems as GBDT, with high calculation and storage costs. In order to avoid the defects of the above two methods and ensure the accuracy, Microsoft opened the LightGBM method in 2017. This method draws on the advantages of XGBoost, such as the second-order Taylor expansion of the objective function, the calculation of the leaf node value, and the expression of the tree complexity. At the same time, the histogram algorithm and the growth strategy of Leaf-wise are added to reduce the amount of calculation and memory occupation.

Histogram algorithm integrates large-scale data into the histogram. Firstly, determine how many containers are required for each feature and assign an integer to each container. Then, evenly divide the range of floating-point numbers into several intervals, and the number of intervals is equal to the number of containers. The sample data belonging to the container is updated to the value of the container. Finally, it is represented by histogram. The histogram algorithm does not need to store the pre-ordering results. Only the discrete values of the feature need to be stored in 8-bit integer, which can reduce the memory consumption by 7/8. In other words, XGBoost needs to use 32-bit floating-point numbers to store feature values and 32-bit integer to store indexes, while LightGBM only needs 32-bit to store histograms.

The schematic diagram of Leaf-wise growth strategy is shown in Fig. 6. The strategy finds the leaf with the largest splitting gain from all the current leaves each time, and then splits, and so on. With the same splitting times, the error of Leaf-wise strategy is

lower than that of XGBoost's Level-wise strategy. LightGBM adds a maximum depth limit to the Leaf-wise strategy to ensure high efficiency and prevent over fitting.

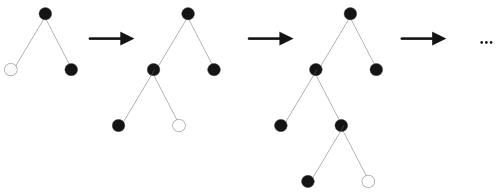

Fig. 6. Schematic diagram of Leaf-wise growth strategy

In this paper, the decision tree model based on LightGBM is used to select the features from the data set of 10000*600. 70 feature points are extracted from each energy curve to construct a 10000*70 feature subset. The algorithm flow is shown in Fig. 7.

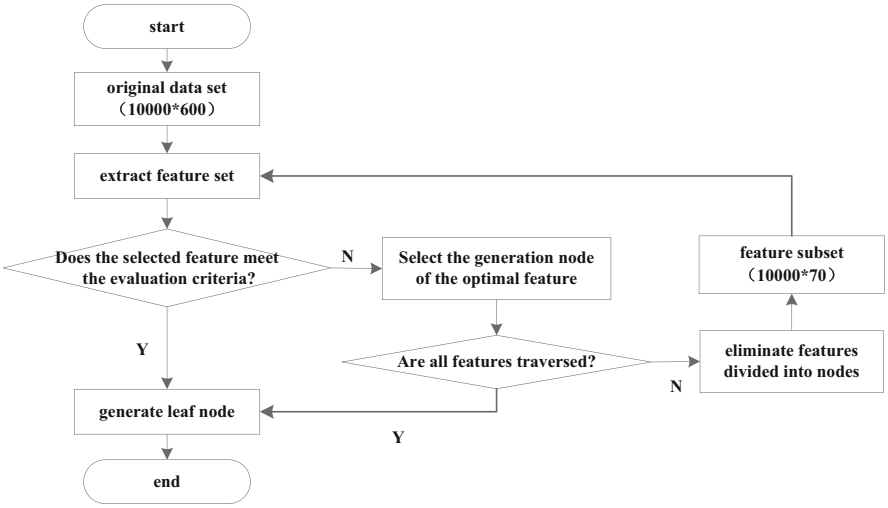

Fig. 7. The flow of feature selection algorithm based on decision tree

The more important the feature is, the better the effect of increasing the node purity is. Function feature_importance in LightGBM model is used to rank the importance of each feature after training. The total gain of the selected feature during the splitting process is shown in Fig. 8. The used times of the selected feature in the model is shown in Fig. 9.

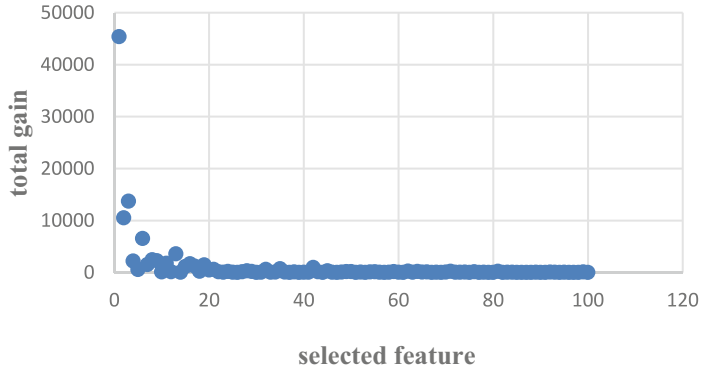

Fig. 8. Total gain of the selected feature during the splitting process

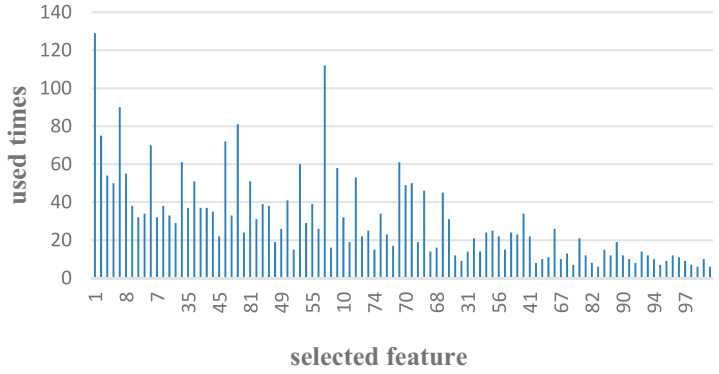

Fig. 9. Used times of the selected feature in the model

In order to avoid over fitting during the algorithm operation, function SelectFromModel in Library sklearn is used. The Logistic Regression with L2 penalty is used as the basic model for training, and then the features are scored and sorted.

3.3 Feature Selection Algorithm Based on RFE

This paper uses Linear Regression model as the base model of the RFE (Recursive Feature Elimination)algorithm. Its optimization goal is to minimize the error of all the training samples. The model can usually be expressed as Formula (5).

$$C = a_0 + a_1 F_1 + \ldots + a_k F_k \tag{5}$$

where C represents the predicted variable value, k represents the number of the selected features, F_i represents the selected i-th feature, and $a_i (i = 0, 1, 2, ..., k)$ represents the regression coefficient.

In the process of searching variables, the stepwise selection method can be adopted, which is improved from the optimal subset method. Stepwise selection method can be

divided into forward selection, backward selection and two-way selection. This paper adopts the forward stepwise selection method.

In the forward selection method, it is assumed that there is only one constant term in the regression equation, and then the introduced variables are selected one by one. In this way, all the independent variables are introduced into the regression equation in turn [10]. Next, analyze the relationship between the introduced variables and the dependent variables in each case. From the set $F = \{x_1, x_2, \ldots, x_m\}$ select a variable with the best linear correlation with the dependent variable as x_{i1}, and then select a variable x_{i2} among the $m - 1$ independent variables to make the combination $D = \{x_{i1}, x_{i2}\}$ has the best regression effect; Select a variable x_{i3} from the remaining $m - 2$ independent variables and put it into D to make the new combination $D_1 = \{x_{i1}, x_{i2}, x_{i3}\}$ has the best regression effect. This step is repeated until the obtained regression equation is optimal, and the subset D_n is the selected optimal feature subset.

The flow of feature selection algorithm based on RFE is shown in Fig. 10.

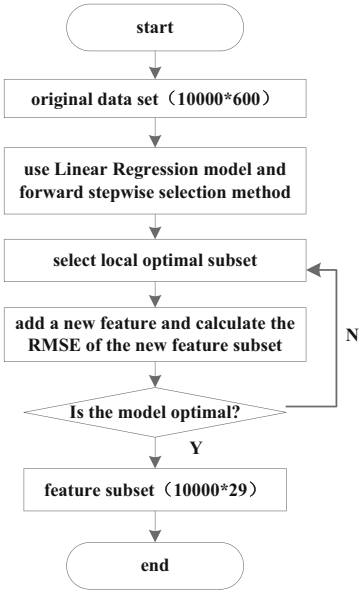

Fig. 10. The flow of feature selection algorithm based on RFE

The stepwise selection method has a large amount of calculation. The commonly used indicators include AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and RMSE (Root Mean Square Error). In this paper, RMSE is selected as the judgment index of feature selection, and its calculation method is shown in Formula (6).

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2} \tag{6}$$

where $\widehat{y_i} = \{\widehat{y_1}, \widehat{y_2}, ..., \widehat{y_n}\}$ is the predicted value and $y_i = \{y_1, y_2, ..., y_n\}$ is the real value. The greater the deviation between the predicted value and the real value, the greater the value. When the predicted value is completely consistent with the real value, the value of *RMSE* is 0, reaching the ideal model.

In the feature selection algorithm based on RFE, RMSE is used to select the local optimal set from the original data set and remove the selected feature values. Then gradually add new features to the local optimal set, and use the returned RMSE value to judge whether the model performance will be improved after adding new features. After the judgment, add the features that are beneficial to the model performance to the local optimal set. At the same time, whether the judged feature is beneficial to the model or not, it will be removed from the remaining features.

The training times of the model and the returned RMSE value are shown in Fig. 11. After selecting the local optimal solution, 19 eigenvalues in favor of the model are introduced in turn. The root mean square error is within a reasonable range and has a significant decrease.

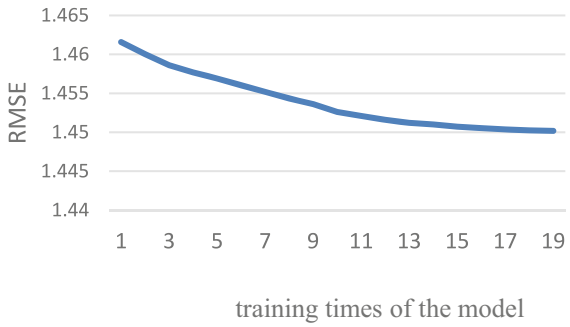

Fig. 11. The training times of the model and the returned RMSE value

3.4 Performance Test of Feature Selection Algorithm

After the feature subset is selected by the feature selection algorithm, it is usually necessary to evaluate the classification of the feature subset to judge the advantages and disadvantages of the feature selection algorithm. This paper adopts the hold-out method to divide the original data set into two mutually exclusive parts: test set and training set. The training set is fully learned through algorithms to obtain the classification model. Then each sample in the test set is studied using the classification model to obtain the class label and calculate the classification accuracy. In this paper, two evaluation criteria, classification accuracy and F1 measure, are used to evaluate the feature subsets selected by the three feature selection algorithms. At the same time, an external learning algorithm is introduced to test the selected feature subset. Since a single classifier will bias to some algorithms and lead to over fitting, two classical classifiers are selected, namely SVM (Support Vector Machine) and KNN (K-Nearest Neighbor).

1. Comparison of classification accuracy

In the model evaluation, classification accuracy is the most used indicator, which refers to the proportion of correctly classified samples to all samples in the test set reserved from the original data set. The calculation method is shown in Formula (7).

$$A = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

In formula (7), TP indicates that the model is predicted as True Positive, that is, the real class of the sample is Positive and the model recognition result is also Positive. FN indicates False Negative, that is, the real class of the sample is Positive, but is recognized as Negative. FP indicates False Positive, that is, the real class is Negative, but is recognized as Positive. TN indicates True Negative, that is, the real class is Negative, and the model recognition result is also Negative.

In this paper, three feature selection algorithms are used to obtain feature sets. The classification accuracy of these sets based on SVM and KNN is shown in Fig. 12.

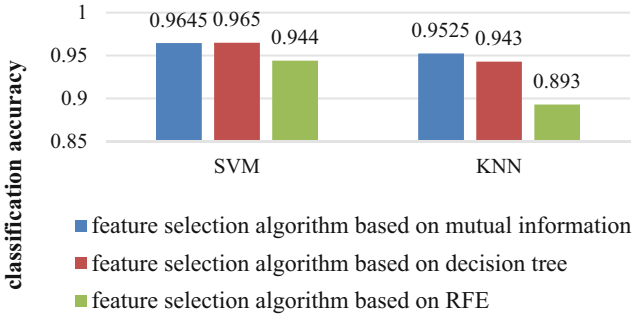

Fig. 12. Classification accuracy of three feature selection algorithms based on SVM and KNN

As can be seen from Fig. 12, in SVM classifier, the accuracy of the three feature selection methods is more than 90%. Among them, the accuracy of the feature selection algorithm based on mutual information is almost the same as that based on decision tree, both reaching more than 96%. Although the accuracy of feature selection algorithm based on RFE is not as good as the former two, it also reaches 94%.

Among KNN classifiers, the feature selection algorithm based on mutual information has the highest classification accuracy, reaching 95%. Although the accuracy of the feature selection algorithm based on decision tree is lower than the former, the difference is less than 1%. In contrast, the feature selection algorithm based on RFE has a lower accuracy, only 89%.

Generally speaking, in terms of classification accuracy, feature selection algorithms based on mutual information and decision tree have good performance, and feature selection algorithms based on RFE have good performance for some specific machine learning algorithms.

2. Comparison of F1 measure

Since a single evaluation index cannot completely judge the merits of an algorithm, F1 measure is also adopted in this paper to make the test results more comprehensive. F1 measure is the harmonic average of precision and recall, and its calculation method is shown in Formula (8).

$$F1 = \frac{2 \times P \times R}{P + R} \tag{8}$$

where, P is the precision and R is the recall, which are respectively expressed in Formula (9) and Formula (10).

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

The value range of F1 measure is [0, 1], where "1" represents that the output of the classification algorithm is the best, and "0" represents that the output of the classification algorithm is the worst.

In Python, f1_score in sklearn.metrics library is used to calculate F1 measure, and the result is shown in Fig. 13.

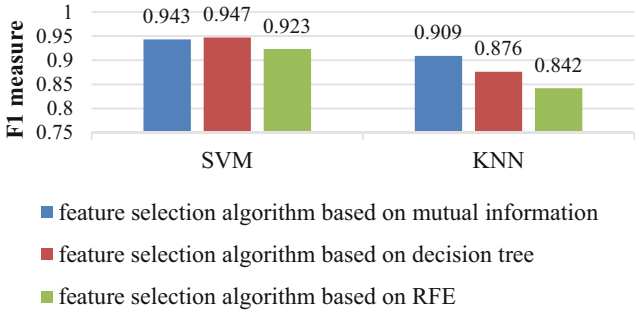

Fig. 13. F1 measure of three feature selection algorithms based on SVM and KNN

It can be seen from the above that the larger the F1 measure, the better the classification performance. As can be seen from Fig. 13, In SVM classifiers, the feature selection algorithm based on decision tree has the largest F1 measure, and the feature selection algorithm based on RFE has the lowest F1 measure. In KNN classifier, the feature selection algorithm based on mutual information has the largest F1 measure, and the feature selection algorithm based on RFE has the lowest F1 measure. Therefore, in terms of F1 measure, the feature selection algorithm based on mutual information has better classification performance.

The three feature selection algorithms are evaluated with different evaluation indexes. The results show that the feature selection algorithm based on mutual information and the

feature selection algorithm based on decision tree have good performance for different classifiers. The feature selection algorithm based on RFE can get good performance in specific classifiers.

The feature selection algorithm based on decision tree is far superior to the other two algorithms in computing speed and memory space. The feature selection algorithm based on mutual information needs to analyze the feature-dependent redundancy ratio, which increases the amount of computation and memory, but improves the performance of the selected feature subset. The feature selection algorithm based on RFE is affected by the internal encapsulation algorithm, which takes a long time and uses a large memory space. However, the feature subset dimension of this algorithm is far lower than the other two, and the subset also has better performance in specific machine learning algorithms.

4 Summery

Based on the research background of energy analysis attack, this paper first introduces three feature selection methods of filtering mode, wrapper mode and embedded mode, and analyzes their advantages and disadvantages. Secondly, three energy curve feature selection algorithms are studied and implemented: dynamic feature selection algorithm based on mutual information, feature selection algorithm based on decision tree and feature selection algorithm based on RFE. The feature points in each energy curve that are highly related to the Hamming Weight of the first S-box output are successfully selected from the original data set. Finally, SVM and KNN are selected as classifiers, and classification accuracy and F1 measure are used as evaluation criteria to evaluate the three feature subsets, so as to judge the merits of the algorithms.

From the test, the feature subsets selected by the three feature selection algorithms have good performance and can be used for energy analysis attacks. The feature selection algorithm based on decision tree has high efficiency, less memory, good accuracy and excellent overall performance. The dynamic feature selection algorithm based on mutual information has high accuracy. Although the operation efficiency is not as high as that of decision tree feature selection algorithm, it is also relatively fast and within an acceptable range. The feature selection algorithm based on RFE belongs to the embedded mode, which has its own shortcomings of low efficiency and time-consuming. It is not good for processing high-dimensional data features such as energy curve, but it still has good performance in low dimensional data.

References

1. Kira, K., Rendell, L.A.: The feature selection problem: traditional methods and a new algorithm. In: Tenth National Conference on Artificial Intelligence,pp.129–134. AAAI Press (1992)
2. John, G.: Irrelevant features and the subset selection problem. In: 11th International Conference on Machine Learning (ICML-94), pp.121–129. Morgan Kaufmann (1994)
3. Koller, D., Sahami, M.: Toward optimal feature selection. In: 13th International Conference on Machine Learning (ICML-96), pp.284–292. Morgan Kaufmann (1996)

4. Kocher, P.C.: Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 104–113. Springer, Heidelberg (1996). https://doi.org/10.1007/3-540-68697-5_9
5. Chari, S., Rao, J.R., Rohatgi, P.: Template attacks. In: Kaliski, B.S., Koç, çK., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 13–28. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36400-5_3
6. Gierlichs, B., Lemke-Rust, K., Paar, C.: Templates vs. stochastic methods. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 15–29. Springer, Heidelberg (2006). https://doi.org/10.1007/11894063_2
7. Archambeau, C., Peeters, E., Standaert, F.-X., Quisquater, J. -J.: Template attacks in principal subspaces. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 1–14. Springer, Heidelberg (2006). https://doi.org/10.1007/11894063_1
8. Czarnowski, I., Jędrzejowicz, P.: Data reduction and stacking for imbalanced data classification. J. Intell. Fuzzy Syst. **37**(6), 7239–7249 (2019)
9. Khosla, K., Jha, I.P., Kumar, A., Kumar, V.: Local-Topology-Based Scaling for Distance Preserving Dimension Reduction Method to Improve Classification of Biomedical Data-Sets. Algorithms **13**(8), 192–203 (2020)
10. Shi, Q., Pan, F., Long, F.: A review of feature selection methods. Microelectron. Comput. **39**(3), 1–8 (2022)
11. Visalakshi, S., Radha, V.: A literature review of feature selection techniques and applications: review of feature selection in data mining. In: 2014 IEEE International Conference on Computational Intelligence and Computing Research, pp. 1–6. IEEE, Coimbatore, India (2014)
12. Aboudi, N.E., Benhlima, L.: Review on wrapper feature selection approaches. In: International Conference on Engineering & MIS (ICEMIS), pp. 1–5. IEEE, Agadir, Morocco (2016)