Individual Identification Method for Frequency-Hopping Radios Based on CMFS-MIC Feature Selection (Postprint)

Authors: Yang Yinsong, Guo Ying, Li Hongguang, Sui Ping, Yu Xinyong

Date: 2018-10-11T00:00:00+00:00

Abstract

To address the problems of high computational complexity and low recognition accuracy in radio identification arising from redundant features in the subtle feature sets of frequency-hopping radios, this paper proposes an individual identification method for frequency-hopping radios based on CMFS-MIC feature selection. Initially, the subtle feature sets are computed for each collected frequency-hopping radio signal sample. Subsequently, correlation information entropy is utilized to measure the combined effect of feature subsets, and each feature is ranked in descending order by comprehensively considering both the correlation and redundancy relationships among features. Building upon this, an approximate Markov blanket method based on maximum information coefficient metric is employed to eliminate redundant features, thereby achieving optimization and dimensionality reduction of the feature subset. Finally, a voting ensemble classifier is designed to identify signals from four frequency-hopping radios. Simulation results indicate that the proposed algorithm yields a higher sorting and recognition rate.

Full Text

0 Introduction

Frequency hopping communication refers to a communication method where both parties, under the control of pseudo-random codes and synchronization algorithms, synchronously hop the radio frequency across a predetermined frequency set according to the pseudo-random code. This communication mode is widely used in military communications due to its excellent anti-jamming capability, low probability of interception, and strong networking ability [?]. As the electromagnetic environment becomes increasingly complex, frequency hopping radio individual identification and network sorting technology have become critical issues to be addressed in the field of communication countermeasures. Simultaneously, researching identification technology for frequency hopping radios and sorting each radio from numerous frequency hopping communication networks is of significant importance for communication reconnaissance, friend-or-foe identification, and implementing electronic attack operations.

Existing frequency hopping signal network sorting and radio individual identification technologies can be mainly divided into two categories: parameter estimation-based network sorting technology and blind source separation-based network sorting technology. Parameter estimation-based network sorting technology primarily utilizes parameter information such as hopping period, Direction of Arrival (DOA), power, and signal temporal correlation to achieve frequency hopping signal identification and classification. Reference [?] employs the MUSIC algorithm to estimate the DOA of frequency hopping signals and uses it for cluster analysis to achieve signal sorting. Reference [?] utilizes estimated hopping period, DOA information, and power, combined with an improved K-means algorithm and statistical histogram methods for frequency hopping signal clustering and sorting. The main problem with these algorithms is the small number of estimable parameters, low estimation accuracy, high dependence of sorting accuracy on feature parameter estimation precision, and rapidly deteriorating performance in complex battlefield electromagnetic environments. Blind source separation-based network sorting technology mainly adopts a "two-step" approach: first estimating the mixing matrix, then utilizing signal sparsity to complete source signal and noise separation under the condition of a known mixing matrix. The primary issues are low correct separation rate of source signals and difficulty in estimating the mixing matrix under underdetermined conditions. In practical electromagnetic environments, limited by system complexity and platform constraints, the number of received mixed signals often exceeds the number of antenna array elements, making this type of method highly limited for practical application.

Considering that even for any two radios from completely identical manufacturing processes, there still exist detectable and reproducible subtle feature differences in signal representation that do not affect information transmission, individual radio sorting can also be achieved by extracting these subtle feature differences. In recent years, some research has begun focusing on individual identification through extracting subtle radio features. Reference [?] proposed extracting the information dimension and box dimension of frequency hopping signal instantaneous envelopes and using neural network construction for classification and identification. Since bispectrum can effectively suppress Gaussian noise, it is widely used to characterize signal subtle features. In reference [?], Xu Shuhua et al. proposed extracting Signal Integrated Bispectrum (SIB) features, combined with Principal Component Analysis (PCA) for feature dimensionality reduction, and finally achieved over 90% resolution on real off-site collected datasets using SVM classification algorithms, though with low robustness in identification effectiveness. Lei Yingke et al. [?, ?] built upon reference [?], established a maximum entropy model to remove non-Gaussian noise effects, and introduced a Collaborative Representation Classifier (CRC) in the classification stage, enabling individual identification of different FM radios of the same model, manufacturer, and operating mode with more robust, rapid, and efficient identification results. Reference [?] considered that different features have varying importance for radio identification rate, introduced a neighborhood rough set data reduction algorithm to evaluate the importance of each feature in classification, and designed a weighted voting combination classifier to improve classification accuracy, but did not reduce computational complexity. Reference [?] demonstrated through simulation that simply piecing together all features does not necessarily form the optimal feature set; more features do not guarantee the best identification rate, and proposed using PCA to select features with high importance for classification, but this method ignores redundancy between features.

In summary, extracting multiple subtle features of radio signals, selecting the feature subset with the best identification performance, and combining machine learning classifier design theory to achieve radio sorting and identification has gradually become a new research trend. This paper proposes an improved correlation information entropy-based feature selection method to optimize the selection of subtle feature sets for frequency hopping signals and constructs a voting combination classifier to achieve frequency hopping radio sorting. To verify the effectiveness of the proposed algorithm, experiments are conducted on both UCI partial datasets and frequency hopping signal feature sets, using sorting accuracy as the evaluation criterion.

1 Preliminary Knowledge

Feature selection refers to the process of eliminating irrelevant or redundant features from an existing set of M features and selecting N most effective features as a new feature subset ($M > N$), such that the new feature subset can significantly reduce computational complexity while providing classification performance approximate to or better than that of the original dataset. References [?, ?] propose mutual information-based feature selection methods that consider the correlation between features and classes but do not fully account for redundancy between features, introducing considerable redundant information, and the algorithms require specifying the number of features to select. References [?, ?] propose using symmetric uncertainty as a correlation evaluation criterion, considering both feature-class and feature-feature correlations to delete redundant features, but the algorithm does not comprehensively consider the combination effect of feature subsets. References [?, ?] apply correlation information entropy theory from the data fusion domain to feature selection, considering the combination effect of feature sets, i.e., multivariate relationships between different features, but may result in unreasonably large or small dimensions of the optimal feature subset. This paper proposes an improved correlation information entropy-based feature selection method to optimize the selection of subtle feature sets for frequency hopping signals. First, two basic concepts are introduced.

1.1 Correlation Information Entropy

Assuming a multivariate system has N variables, each variable has M different time moments, then the multivariate time series matrix constructed by this system is:

$$P = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix}$$

where $P \in \mathbb{R}^{N\times M}$ is a real matrix. After centering and standardization, we obtain Q, and calculate its correlation matrix:

$$R = Q^T Q$$

For N variables, the correlation matrix R can be expressed as:

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ R_{21} & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{N1} & R_{N2} & \cdots & R_{NN} \end{bmatrix} = \begin{bmatrix} I & R_{12} \\ R_{21} & I \end{bmatrix}$$

where I is the auto-correlation matrix and R_{12} is the cross-correlation matrix representing the overlapping information between variables. Assuming the eigenvalues of matrix R are $\lambda_1, \lambda_2, \dots, \lambda_N$. According to information theory calculation principles, the correlation information entropy is defined as [**?**]:

$$H(R) = -\frac{1}{N}\sum_{i=1}^{N}\log\lambda_i$$

Since the eigenvalues of the identity matrix I are all 1, then:

$$H(I) = -\frac{1}{N}\sum_{i=1}^{N}\log 1 = 0$$

Combining the eigenvalue conditions and equations (4)-(5), we know that when $\lambda_i = 1$ for $i = 1, 2, \dots, N$, it indicates that the variables are mutually independent, providing completely different information with no overlapping information in the system, and the correlation information entropy $H(R) = 0$. When $\lambda_i \neq 1$, the system contains overlapping information.

1.2 Maximum Information Coefficient

Reshef et al. proposed a new information theory-based metric in Science in 2011—the Maximum Information Coefficient (MIC). MIC can measure not only linear and nonlinear relationships between variables in large datasets but is also effective for non-functional dependencies that cannot be represented by a single function [?].

MIC is primarily calculated using mutual information and grid partitioning methods. Given variables $X = \{x_i, i = 1, 2, \dots, n\}$ and $Y = \{y_i, i = 1, 2, \dots, n\}$ with n sample pairs, for the finite sample pair set $D = \{(x_i, y_i), i = 1, 2, \dots, n\}$, a grid G partitions the value domain of variable X into x segments and variable Y into y segments, forming an $x \times y$ grid. For each grid partition, mutual information $I(D|G)$ is calculated. There can be multiple grid partitioning methods, and the maximum value $I^*(D, x, y)$ is selected as the mutual information value for D under partition (x, y). The formula for maximum mutual information under partition (x, y) for data D is:

$$I^*(D, x, y) = \max I(D|G)$$

The characteristic matrix $M(D)_{x,y}$ consists of the maximum mutual information values obtained under different partitions. The Maximum Information Coefficient is then defined as:

$$MIC(D) = \max_{xy<B(n)} \frac{I^*(D, x, y)}{\log \min\{x, y\}}$$

where $B(n)$ is the upper limit of grid partition xy. Reference [?] shows through extensive experiments that the best results are achieved when $B(n) = n^{0.6}$.

2 Proposed Algorithm

2.1 Improved Correlation Information Entropy-Based Feature Selection Algorithm

Let the original dataset D contain N samples, each sample consisting of M features forming set $X = \{f_1, f_2, \dots, f_M\}$. The target set C contains p different categories $\{c_1, c_2, \dots, c_p\}$. Feature selection aims to select an optimal feature subset S from the M-dimensional feature set, where S contains N_S features ($N_S < M$). This feature subset S provides approximate or better recognition capability during classification compared to the original set X.

The CMFS algorithm based on correlation information entropy is briefly described as follows [?]:

a) Calculate the mutual information $I(f_i; c)$ between each feature and its corresponding class, and construct the multivariate real matrix P (where $H(\cdot)$ is joint entropy);

b) Center and standardize the real matrix P to obtain Q, and calculate the feature correlation matrix $R = Q^T Q$;

c) Calculate the eigenvalue vector of matrix R, and compute $H(R)$ and $H(I)$ for all eigenvalues according to equations (4) and (5);

d) For each feature f_i, calculate the increase in redundant information after losing feature f_i based on $H(R_{\text{all}}) - H(R_{\text{miss}_i})$, where $H(R_{\text{all}})$ is the correlation information entropy under all features, and $H(R_{\text{miss}_i})$ is the correlation information entropy after losing feature f_i;

e) Select the first feature $f_{(1)}$ according to $\arg\max_{f_i \in X}(\text{Info}(i))$, where $\text{Info}(i) = H(R_{\text{all}}) - H(R_{\text{miss}_i})$;

f) For each remaining feature f_i, add f_i to set S, the expanded correlation matrix is R_{add}, calculate the correlation information entropy after expansion: $H(R_{\text{add}})$, and iteratively sort according to $\arg\max_{f_i \in X-S}(H(R_{\text{add}}))$.

The CMFS algorithm ensures through feature ranking that the feature subset formed by the top l dimensions is low-redundancy and has maximum correlation information entropy, but cannot adaptively select the feature subset and requires manual setting of dimension l. Reference [?] introduces a threshold η for adaptive control of subset size, called the CMFS-n algorithm. The difference from CMFS lies in step 6: when $H(R_{\text{add}}) - H(R_{\text{all}}) < \eta$, feature f_i is added to set S; when the added redundancy information exceeds the limit, the algorithm terminates iteration. The CMFS-n algorithm achieves adaptive control of feature subsets, but the fundamental problem is how to select the value of η. When η is small, almost no redundant information is allowed, and the feature subset size nearly equals the original set size, failing to achieve dimensionality reduction. When η is large, less redundant information is allowed, leaving too few features and affecting sorting recognition rate. Reference [?] determines the threshold η through extensive experiments considering both classification accuracy and computational efficiency, but can only provide a range of good performance, and the value differs for different datasets. This method of finding the optimal threshold η through extensive experiments seriously increases computational complexity.

Reference [?] proposes a feature selection method based on the Maximum Information Coefficient metric and approximate Markov blanket. First, the FCBF algorithm uses symmetric uncertainty to measure feature ranking and delete weakly relevant features, then applies the approximate Markov blanket to delete redundant information and obtain the optimal feature subset. Inspired by this, this paper also adopts the approximate Markov blanket method to delete redundant information and achieve adaptive feature subset selection.

Definition 1 (Approximate Markov Blanket). For two features f_i and f_j, f_j is an approximate Markov blanket of f_i if:

$$MIC(f_i, c) > MIC(f_j, c) \quad \text{and} \quad MIC(f_i, f_j) > MIC(f_j, c)$$

Definition 2 (Primary Element). A feature f_i is a primary element if and only if no element in feature set X is an approximate Markov blanket of f_i. For any dataset, the feature subset composed of all primary elements is the optimal feature subset.

Based on the above analysis, this paper proposes an improved correlation information entropy-based feature selection algorithm, abbreviated as CMFS-MIC, which first uses the CMFS algorithm for feature value ranking, then uses the Maximum Information Coefficient-based approximate Markov blanket to remove redundant features. The specific algorithm steps are as follows:

Input: Dataset D, feature set X, class information C
Output: Optimal feature subset S_{opt}

a)-e) Same as steps a)-e) of the CMFS algorithm;

f) Same calculation method as step f) of the CMFS algorithm, but perform feature ranking for all M features and place them in S;

g) Select the first feature x_1 in S as a primary element, add it to S_{opt}, and find the feature subset $\{x_j\}$ that has x_1 as a Markov blanket. Delete the redundant feature subset $\{x_j\}$ from S to obtain a new S;

h) Repeat step g) until $S = \emptyset$, then output S_{opt}.

2.2 Voting Combination Classifier Design

In classifier design, constructing a combination classifier can achieve better classification performance by fusing decisions from multiple classifiers and making a second judgment on the outputs of different classifiers through a combiner.

In frequency hopping signal feature sets, compared with three higher-order spectral features—Radial Integrated Bispectrum (RIB), Axially Integrated Bispectrum (AIB), and Circularly Integrated Bispectrum (CIB)—the SIB feature has better advantages in time-shift invariance, scale variability, and phase preservation, and can best reflect individual differences among radiation sources in classification. Therefore, this paper focuses only on SIB features in bispectrum features. Additionally, simulation experiments in references [**?**, **?**] show that high-dimensional SIB alone has good classification effects, but when combined with other low-dimensional features to form a new feature vector, the sorting recognition rate 反而 decreases. Therefore, the combination classifier designed in this paper inputs SIB feature sets and low-dimensional feature sets into two CRC classifiers [**?**] and two SVM classifiers separately, and finally performs voting combination judgment on the results of the four classifiers. The voting combination classifier is shown in [Figure 1: see original paper].

2.3 Overall Algorithm Flow

Based on the above theoretical analysis of the CMFS-MIC feature selection algorithm and weighted combination classifier, this paper proposes a frequency hopping radio sorting method based on CMFS-MIC feature selection. The overall algorithm flow framework is shown in [Figure 2: see original paper].

3 Experimental Results and Analysis

3.1 Experiment 1

To verify the effectiveness of the proposed feature selection algorithm CMFS-MIC, several commonly used datasets from the UCI Machine Learning repository [?] were selected (as shown in) for feature selection, and compared with classical feature selection algorithms FCBF and CMFS-n. Two classifiers were used for comparative experiments. During the experiment, the original data was randomly divided into 50% training samples and 50% test samples. To improve experimental accuracy, the experiment was repeated 10 times and the average was taken as the final result.

shows the UCI datasets: Dermat, Spambase, Arrhyth, and Colon.

and respectively show the feature dimensionality reduction capability of FCBF, CMFS-n, and CMFS-MIC on the datasets in , and the classification accuracy of the optimal feature subsets selected by the three different methods on the test sets. The threshold η for the CMFS-n algorithm was selected through multiple experimental simulations considering both classification success rate and feature dimensionality reduction. "Full" represents the complete data set, and "Average" represents the mean value.

From , it can be seen that all three algorithms can achieve dimensionality reduction, but their capabilities differ. The proposed CMFS-MIC feature selection algorithm obtained fewer feature dimensions in these four datasets, demonstrating superior dimensionality reduction capability compared to the other two algorithms. From , compared with the FCBF algorithm, CMFS-MIC can achieve higher recognition rates with smaller data feature dimensions. Compared with the CMFS-n algorithm, CMFS-MIC has respective advantages in classification performance, but CMFS-n requires multiple experiments to determine the threshold η, and different values significantly affect classification results and feature dimension selection, resulting in poor robustness and time-consuming implementation. Additionally, after feature selection, the accuracy of different datasets on different classifiers has improved to some extent. On the same dataset, the CRC algorithm achieves higher classification accuracy than SVM.

3.2 Experiment 2

This experiment first collected data from 4 shortwave frequency hopping radios of the same model. The data collection instrument was a Tektronix DPO 7254 digital phosphor oscilloscope with a sampling rate up to 40 GHz/s. Each radio collected 100 samples, with 50 used for training and 50 for testing. The experimental platform was a Lenovo T440S PC with 12GB memory, Intel(R) CoreTM i7-4600U CPU @2.10 GHz processor, Windows 7 64-bit operating system, and MATLAB R2014a simulation software.

The collected signal samples from the 4 frequency hopping radios were preprocessed and feature sets were calculated. Let the obtained feature set be $F = \{f_1, f_2, \ldots, f_{256}\}$, where: - f_1: Spectral symmetry mean - f_2: Spectral symmetry variance
- f_3: Rayleigh entropy - f_4: Information dimension - f_5: Box dimension - f_6: LZC - f_7: Envelope kurtosis - f_8: High-order J feature mean - f_9: High-order J feature variance - f_{10} to f_{256}: SIB features

Class information $C = \{1, 2, 3, 4\}$. After feature selection by CMFS-MIC, the optimal feature subsets $F_1^* = \{f_1, f_3, f_5, f_6, f_8\}$ and $F_2^* = \{f_{127}, f_{221}, f_{10}, f_{1010}\}$ are obtained, constructing feature sets $F_3 = \{F_1, F_2\}$ and $F_4 = \{F_1^*, F_2^*\}$. Since the CRC classifier achieves better results than SVM, this paper only presents the recognition rates of the CRC classifier under different feature sets compared with the proposed algorithm, as shown in .

Analysis of the results in shows: Comparing F_1 with F_1^* and F_2 with F_2^* indicates that whether conventional low-dimensional feature sets or high-dimensional integrated bispectrum feature sets, the classification recognition rate after feature selection is slightly lower than that of the original feature set, but the difference is small, demonstrating that the feature set obtained through CMFS-MIC still maintains approximate sorting capability to the original feature set while significantly reducing dimensionality. Comparing F_3 and F_4 shows that the recognition rate of the complete feature set is not as good as that of the low-dimensional feature set obtained after feature selection. The recognition rate results of the proposed algorithm show that using the voting combination classifier method significantly outperforms single CRC classifier classification performance.

4 Conclusion

This paper proposes a frequency hopping radio network sorting method based on CMFS-MIC feature selection. The method considers both the correlation between features and classes and the redundancy between features to rank the feature set, and uses the approximate Markov blanket method to delete redundant and irrelevant features to obtain the optimal feature subset. In the classification and identification stage, a voting combination classifier is designed based on SVM and CRC classifiers. Experimental results verify the effectiveness of the proposed feature selection algorithm CMFS-MIC and demonstrate that the proposed algorithm improves the recognition rate of radio sorting. However, the limitation is that the feature selection process is computationally complex, resulting in still high overall computational complexity for a single experiment, which needs improvement in future research.

References

[1] Sha Zhichao, Liu Zhangmeng, Huang Zhitao, et al. Online hop timing detection and frequency estimation of multiple FH signals[J]. ETRI Journal, 2013, 35(5):748–756.

[2] Eric M, Dukic M L, Obradovic M. Frequency hopping signal separation by spatio-frequency analysis based on the music method [C]// Proc of the 6th IEEE International Symposium on Spread Spectrum Techniques and Applications. Piscataway, NJ: IEEE Press, 2000: 78-82.

[3] Chen Lihu, Zhang Eryang, Shen Rongjun. The sorting of frequency hopping signals based on K-means algorithm with optimal initial clustering centers[J]. Journal of National University of Defense Technology, 2009, 31(2):70-75.

[4] Guo Haizhao, Zhang Shunsheng. Application of Sparse Bayesian model in frequency-hopping signal station separation[J]. Journal of Signal Process, 2016, 32(6):733-738.

[5] Gu Chenhui, Wang Lunwen. Individual frequency-hopping radio identification method based on transient characteristics of frequency domain [J]. Journal of Signal Process, 2012, 28(9):1335-1340.

[6] Xu Shuhua, Huang Benxiong, Xu Lina. Identification of individual radio transmitters using SIB/PCA[J]. Journal of Huazhong University of Science Technology: Natural Science Edition, 2008, 36(7):14-17.

[7] Tang Zhe, Lei Yingke. Method of individual communication transmitter identification based on maximum correntropy[J]. Journal of Communication, 2016, 37(12):171-175.

[8] Tang Zhe, Lei Yingke. Radio Transmitter Identification Based on Collaborative Representation[J]. Wireless Personal Communications, 2017, 96(1):1377-1391.

[9] Sun Na. Research on the subtle characteristics of communication stations[D]. Beijing: Beijing University of Posts and Telecommunications, 2010.

[10] Hu Jianshu. Individual character recognition of short wave radios[D]. Guangzhou: South China University of Technology, 2010.

[11] Chandrashekar G, Sahin F. A survey on feature selection methods[J]. Computers & Electrical Engineering, 2014, 40(1):16-28.

[12] Hoque N, Bhattacharyya D K, Kalita J K. MIFS-ND: A mutual information-based feature selection method[J]. Expert Systems with Applications, 2014, 41(14):6371-6385.

[13] Estevez P A, Tesmer M, Perez C A, et al. Normalized Mutual Information Feature Selection[J]. IEEE Trans on Neural Networks, 2009, 20(2):189-201.

[14] Yu Lei, Liu Hu. Efficient Feature Selection Via Analysis of Relevance and Redundancy[J]. Journal of Machine Learning Research, 2004, 5(12):1205-1224.

[15] Duan Hongxiang, Zhang Qiuyu, Zhang Moyi. FCBF algorithm based on normalized mutual information for feature selection[J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2017, 45(1):52-56.

[16] Wang Qiang, Shen Yang, Zhang Yan, et al. Fast quantitative correlation analysis and information deviation analysis for evaluating performances of image fusion techniques[J]. IEEE Trans on Instrumentation & Measurement, 2004, 53(5):1441-1447.

[17] Dong Hongbin, Teng yang, Yang Xue. Feature selection based on measurement of correlation information entropy[J]. Journal of Computer Research and Development, 2016, 53(8):1684-1695.

[18] Sun Guanglu, Song Chao, Liu Jinlai, et al. Feature selection method based on maximum information coefficient and approximate Markov blanket[J]. Acta Automatica Sinica, 2017, 43(5):795-805.

[19] Reshef D N, Reshef Y A, Finucane H K, et al. Detecting Novel Associations in Large Data Sets[J]. Science, 2011, 334(6062):1518.

[20] Lichman M. UCI machine learning repository [EB/OL]. (2015-11-10). http://archive.ics.uci.edu/ml.

Note: Figure translations are in progress. See original paper for figures.