Modulation Recognition Technology of Communication Signals Based on Density Clustering and Sample Reconstruction

Hui Han[1], Xianglong Zhou[2], Xiang Chen[1], Ruowu Wu[1], and Yun Lin[2](✉)

[1] State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang 471003, Henan, China

[2] College of Information and Communication Engineering, Harbin Engineering University, Harbin, China

linyun_phd@hrbeu.edu.cn

Abstract. Modulation recognition is an important part in the field of communication signal processing. In recent years, with the development of modulation recognition technology, various problems have emerged. In this title, we propose an improved recognition framework based on SVM, which extracts the entropy feature of the signal and distinguishes it from the traditional modulation recognition framework. We combine the training set with the test set first, then carry on the density clustering to the whole data set. The data set after the cluster is extracted according to a certain proportion to build a new training set, and the new training set is used to train the SVM. Finally, the data of the test set is modulated by the modulation recognition. Experimental results show that the proposed method improves the recognition rate of traditional SVM framework and enhances the stability of traditional SVM framework.

Keywords: Modulation recognition · Entropy features · Density clustering · Restructure · SVM

1 Introduction

Modulation recognition technology is a key technology in the field of communication processing digital signals. He has been widely used in civil and military applications. As the channel environment is more and more complex, the accuracy of modulation recognition is particularly important.

Traditional modulation recognition technology is constrained by the theoretical framework, and many scholars are devoted to improving the classification and training structure of classifier to improve the accuracy of recognition. With the rise of machine learning, more and more unsupervised methods are widely used [1–5]. Density clustering is an unsupervised classification method. The density clustering algorithm assumes that the clustering structure can be determined by the close degree of the sample distribution, and studies the distribution characteristics between samples from the angle of sample density to achieve the purpose of classification [6, 7]. Density clustering

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 476–484, 2019.
https://doi.org/10.1007/978-3-030-19086-6_53

algorithm is widely used in the processing of complex structure data processing [1]. For large data mining, density clustering also plays an important role [2]. Mass data will appear in the process of generation, which is consistent with the algorithm of density clustering [8]. In addition, density clustering is widely applied in various fields such as biomedical science, computer calculate, economics, and social network [3] and so on.

Entropy is a measure of information uncertainty, which can express the uncertain information of random signals. It can be used as the characteristic of the signal to recognize it [9–14]. Support vector machine (SVM) is a machine learning method based on statistical learning theory. Through training, the support vector machine can automatically find the support vectors that have good distinguishing ability for the classification, thus the classifier can be constructed to distinguish the different categories better [15–17]. At present, support vector machines has a wide range of applications in many fields, such as modulation recognition [15], face recognition [17], etc.

In this paper, a kind of modulation recognition method which combines unsupervised density clustering with SVM is presented. By using data sample and training set to form a new training set, a new training set is trained to achieve better recognition effect. We classify 6 kinds of digital modulation signals: 2ASK, 2FSK, 8FSK, BPSK, QPSK, 16QAM, in different SNR. The algorithm improves the traditional SVM.

2 Related Work

2.1 Entropy Feature Extraction

The concept of entropy was first proposed by Shannon. It is a measure of information uncertainty. Set the event set to. The n-dimensional vector $P = (P_1, P_2, \cdots, P_n)$ is used to represent the probability set of events, and satisfies: $0 \le p_i \le 1$ and $\sum_{i=1}^{n} p_i = 1$, then Shannon entropy can be defined as:

$$H(p) = H(p_1, p_2, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log p_i \tag{1}$$

Assuming that the probability of an event is p_i, the amount of information can be defined as:

$$\Delta I(p_i) = e^{1-p_i} \tag{2}$$

Then the exponential entropy can be defined as:

$$H = \sum_{i=1}^{N} p_i e^{1-p_i} \tag{3}$$

According to the reference [11], in this paper, we extract 6 kinds of entropy characteristics, namely: power spectrum Shannon entropy, wavelet energy spectrum entropy, bispectrum entropy, sample entropy, singular spectrum Shannon entropy, and singular spectral index entropy.

2.2 Density Clustering

As an unsupervised clustering algorithm, density clustering is mainly used to divide the data with the density of data distribution. In general, the distribution of data is presented or sparse or closely distributed according to its characteristics. The data is classified by calculating the core points of the data and dividing the radius of the data categories. In this paper, we use the famous density clustering algorithm: DBSCAN, which partitions data sets through neighborhood, core objects, density radius and other parameters. According to the operation of the algorithm, the following definitions are given [7]:

ε Neighborhood: for $x_j \in D$, its ε neighborhood contains the samples whose concentration is not greater than the sample set.

Core object: if the ε neighborhood of x_j contains at least n samples, it is a core object.

Direct density: if x_j located in the ε neighborhood of x_i and it is the core object, it is called x_j to x_i the density direct.

Density can reach: for x_j and x_i, if there is a sample sequence $p_1, p_2, \ldots, p_n$, among them $p_1 = x_i, p_n = x_j$, and p_{i+1} to p_i is direct density, then it is called x_j to x_i the density can reach.

Density connected: for x_j and x_i, if exist x_k make x_j and x_i are all reached by density, they are called density connected.

Based on these basic concepts, DBSCAN defines "clusters" as the largest set of density connected samples derived from density reachability relationships. The DBSCAN algorithm first selects one of the core objects in the data set as "seeds", and then starts to determine the corresponding clusters. First, find all the core objects according to the given neighborhood parameters, and then use any object as the starting point. Clusters are generated from their densely-reachable samples until all core objects are accessed.

2.3 SVM

SVM is V. Vapnik proposes statistical learning theory. Its basic idea is to perform some kind of nonlinear transformation on the original feature space by defining an appropriate kernel function, mapping the original feature space to a high-dimensional space, and then finding the optimal classification surface in this new space so that the sample is correct. Separation and categorization intervals are the greatest. First, the SVM needs to be trained, and then the trained signal classifier is used to classify the test signals. Figure 1 shows the SVM identification process.

SVM is essentially a two-class classifier. For linearly separable cases, the training sample set is $(x_1, y_1), \cdots, (x_i, y_i)(i = 1, \cdots, n; y_i \in \{-1, 1\})$ using a hyperplane $\omega^T \cdot x + b = 0$ divide $y_1, \cdots, y_n$ error-free into two categories. In practice, it is necessary to separate a sample from multiple samples. This requires the construction of multiple classifiers. The commonly used SVM multivalued classifier construction methods include one-to-many method, one-to-one method, one-time solution method, decision directed acyclic graph method, binary tree-based multi-class support vector machine classification method, and the like. The number of sub-classifiers required by the SVM based on the binary tree structure is small. As the classification progresses, fewer training samples are required and the classification efficiency is high, which is very suitable for the case of many modulation methods.

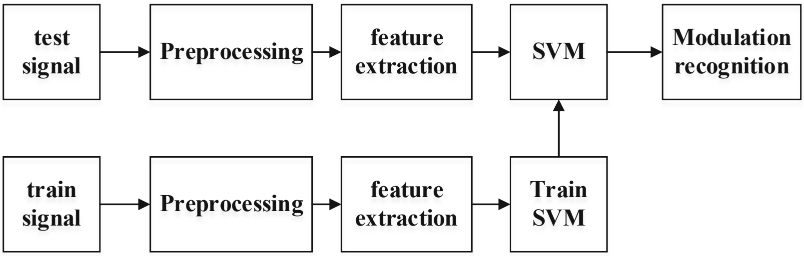


Fig. 1. Identification framework of the SVM

In this paper, a support vector machine based on binary tree structure is selected, and Gaussian kernel functions are used to classify signals.

3 Algorithms and Models

Based on the entropy feature of the signal, this paper adopts the density clustering method to sample and reconstruct the training set, then uses the reconstructed new training set to train the support vector machine, and finally inputs the test set for modulation recognition. Figure 2 shows the specific identification model. We call this model the D-SVM.

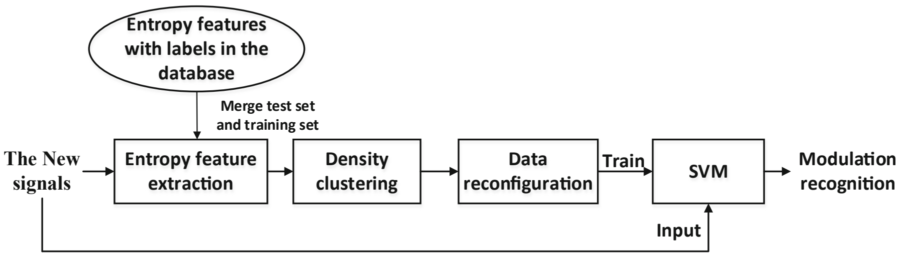


Fig. 2. Modulation recognition model based on density clustering data reconstruction

In order to understand the process of sample refactoring, we demonstrate the basic idea of this method using two dimensional data classification. As shown in Fig. 3, the training samples are represented by solid points, and the samples to be recognized are represented by hollow points. The triangle and the circle represent different categories respectively. As you can see in Fig. 3, when the test set is deviant from the training set, the traditional SVM will have a wrong classification, and our method will effectively repair the deviation to improve the accuracy of the classification.

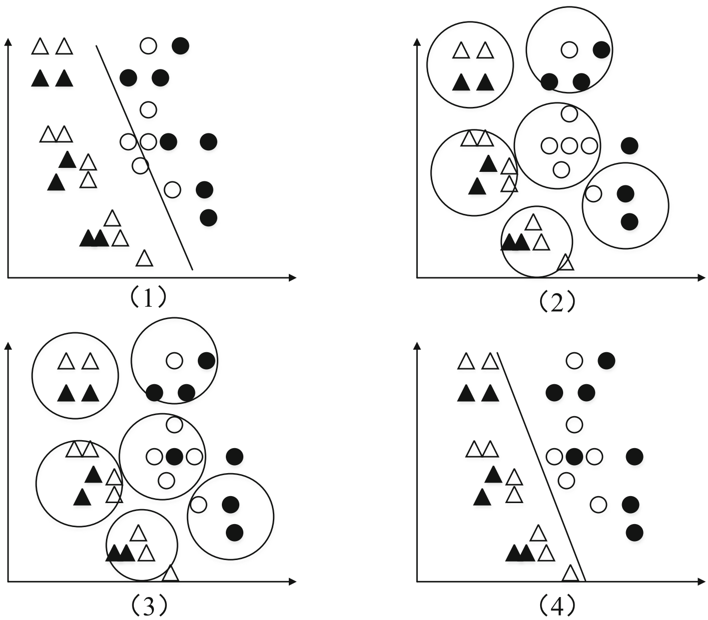


Fig. 3. Schematic diagram of sample reconstruction process

The entropy feature of the received signal is extracted and then integrated with the original entropy feature database. Through density clustering, we can aggregate the data into n subclasses $\{N_1, N_2, \cdots, N_n\}$. Assume that the sample number of the i-th subclass is N_{C_i}, and the number of samples in the train set is N_t and the number of samples in the test set is N_s. In the re-sampling process, we select $N_{C_i}N_t/(N_t+N_s)$ representatives' data from each sub-category to form a new training set. The representative can be calculated by the following formula:

$$Rp(x_i) = \frac{1}{d(x_i, c_k) + \varepsilon} \tag{4}$$

In the formula, $c_k = \frac{1}{N_{c_k}} \sum_{\ell=1}^{N_{c_k}} x_\ell$ represents the cluster center of the k-th subclass of x_i. The constant $\varepsilon = 1.0 \times 10^{-3}$ is to prevent the value of $Rp(x_i)$ from infinity when x_i is just in the center.

The modulation recognition algorithm based on density clustering and SVM is as follows:

Input: train data, test data
Output: Modulation recognition result

Step1: Initialize the cluster sample and merge the training set with the test set $N = N_t \cup N_s$

Step2: Perform density clustering on the merged cluster samples to obtain n subclasses

Step3: According to formula (4), calculate the representative $Rp(x_i)$ of each subcategory N_C, sort the representative data from the largest to the smallest, and select the first $N_{C_i} N_t / (N_t + N_s)$ data to form a new training set.

Step4: Put the test set into the SVM after training the SVM with the new training set to perform modulation recognition

Step5: Output modulation recognition result

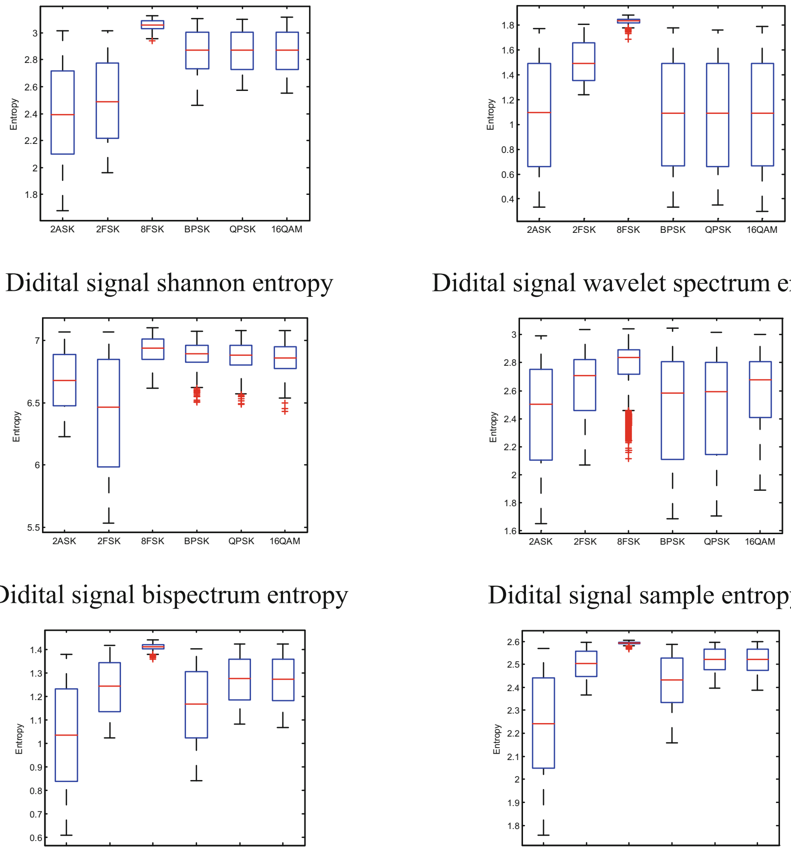
4 Experiment

The experiments in this paper have selected six kinds of digital modulation signals, including 2ASK, 2FSK, 8FSK, BPSK, QPSK and 16QAM. For each signal, the sampling frequency is 16 kHz and the carrier frequency is 4 kHz. The SNR ranges from −5 dB to 10 dB in steps of 1 dB. We extracted six entropy features of the signal, including power spectrum Shannon entropy, wavelet energy spectrum entropy, bispectrum entropy, sample entropy, singular spectrum Shannon entropy, and singular spectrum index entropy. The entire data set is divided into train data and test data. At each SNR, we collected 1200 data points for each entropy feature, each of which has 200 data points.

At first we extracted the entropy features. From Fig. 4 we can see the distribution of entropy features under different SNR. Since the entropy value is some discrete data points, we use the box diagram to better characterize it. We can see from the figure that the sample entropy of the BPSK signal has quite a lot of outliers, which makes it impossible to classify using sample entropy alone. So we adopt the entropy feature fusion method for clustering and modulation recognition.

In this paper, we use the SVM based on the Gaussian kernel function and cross-validation to determine the parameters to classify. We have separately used the traditional SVM architecture and our improved SVM framework based on density clustering reconstruction (D-SVM). The D-SVM recognition framework was tested and the experimental results can be seen form the Fig. 5.

From the curves, we can see that the improved D-SVM recognition framework has approximately 70% recognition rate due to the performance of the traditional method at the SNR of −5 dB. The degree of smoothness is better than the traditional SVM recognition framework, which shows that the stability of the algorithm is better. In the case of a SNR greater than 10 dB, the recognition rates of both are close to 100%. In short, the improvement effect of this algorithm is still ideal.



Didital signal shannon entropy

Didital signal wavelet spectrum entropy

Didital signal bispectrum entropy

Didital signal sample entropy

Didital signal specrtal shannon entropy

Didital signal spectral exponent entropy

Fig. 4. In different signals with mixed signal-to-noise ratios (from −5 dB to 10 dB), the box plot distribution of entropy features includes the maximum, the seventy-fifth percentile, the median, and the twenty-fifth percentile and minimum.

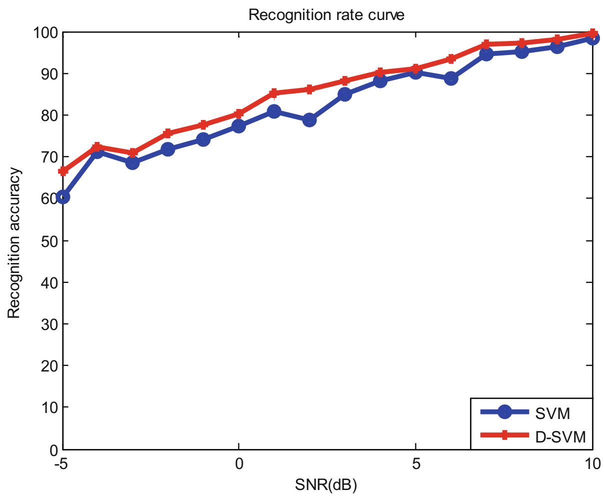


Fig. 5. The recognition rate curves of the two methods at SNR of −5 dB to 10 dB

5 Conclusion

In this paper, we proposes a new method for modulation recognition based on SVM with density clustering data reconstruction. We selected six digital modulation signals, including 2ASK, 2FSK, 8FSK, BPSK, QPSK, and 16QAM, and extracted six entropy features: Power Spectrum Shannon Entropy, Wavelet Energy Spectral Entropy, Bispectrum Entropy, Sample Entropy, Singular Spectrum Shannon Entropy, and Singular Spectrum Exponential Entropy. We fuse these six entropy features, and then density-cluster the entropy features, resample the clustered data set to generate a new training set, and finally train the SVM with the new training set, then the signal Identify. The experimental results show that this method not only improves the accuracy of the traditional support vector machine but also enhances the stability of the recognition framework and has a high application value.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (61771154) and the Fundamental Research Funds for the Central Universities (HEUCFG201830).

Meantime, all the authors declare that there is no conflict of interests regarding the publication of this article.

We gratefully thank of very useful discussions of reviewers.

References

1. Lv, Y., Ma, T., Tang, M., et al.: An efficient and scalable density-based clustering algorithm for datasets with complex structures. Neurocomputing **171**(C), 9–22 (2016)
2. Boonchoo, T., Ao, X., He, Q.: An efficient density-based clustering algorithm for higher-dimensional data (2018)
3. Chen, J., Lin, X., Wu, Y., et al.: Double layered recommendation algorithm based on fast density clustering: case study on Yelp social networks dataset. In: International Workshop on Complex Systems and Networks, pp. 242–252. IEEE (2018)
4. Liu, Z., Yang, J.A., Liu, H., et al.: Transfer learning with fuzzy neighborhood density clustering and re-sampling. J. Sig. Process. **32**, 651–659 (2016)
5. Hu, L., Chan, K.C.: A density-based clustering approach for identifying overlapping protein complexes with functional preferences. BMC Bioinform. **16**(1), 174 (2015)
6. Amini, A., Saboohi, H., Wah, T.Y., et al.: A fast density-based clustering algorithm for real-time Internet of Things stream. Sci. World J. **2014**(1), 926020 (2014)
7. Dai, B.R., Lin, I.C.: Efficient map/reduce-based DBSCAN algorithm with optimized data partition. In: IEEE Fifth International Conference on Cloud Computing, pp. 59–66. IEEE Computer Society (2012)
8. Vedaldi, A., Zisserman, A.: Efficient additive kernels via explicit feature maps. IEEE Trans. Pattern Anal. Mach. Intell. **34**(3), 480–492 (2012)
9. Zhao, Z., Shang, J.: A new method for modulation type recognition based on the time frequency representations. In: International Conference on Signal Processing, vol. 1, pp. 208–211. IEEE (2002)
10. He, Z.Y., Cai, Y.M., Qian, Q.Q.: Study of wavelet entropy theory and its application in electric power system fault detection. In: Proceedings of the CSEE (2005)
11. Liu, T., Guan, Y., Lin, Y.: Research on modulation recognition with ensemble learning. EURASIP J. Wirel. Commun. Netw. **2017**(1), 179 (2017)
12. Pawar, S.U., Doherty, J.F.: Modulation recognition in continuous phase modulation using approximate entropy. IEEE Trans. Inf. Forensics Secur. **6**(3), 843–852 (2011)
13. Kadambe, S., Jiang, Q.: Classification of modulation of signals of interest. In: Digital Signal Processing Workshop. 2004 IEEE Signal Processing Education Workshop, 2004, pp. 226–230. IEEE (2004)
14. Xiao, Y., Wang, H., Xu, W.: Parameter selection of Gaussian kernel for one-class SVM. IEEE Trans. Cybern. **45**(5), 941–953 (2017)
15. Imdad, A., Bres, S., Eglin, V., et al.: Writer identification using steered hermite features and SVM. In: International Conference on Document Analysis and Recognition, pp. 839–843. IEEE (2017)
16. Wu, J., Yang, H.: Linear regression-based efficient SVM learning for large-scale classification. IEEE Trans. Neural Netw. Learn. Syst. **26**(10), 2357–2369 (2017)
17. Osuna, E., Freund, R., Girosi, F.: Training support vector machines: an application to face detection. In: 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Proceeding, pp. 130–136. IEEE (2002)