Expert Systems With Applications 90 (2017) 13–22

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Blind signal modulation recognition through clustering analysis of constellation signature

Gaurav Jajoo[a,*], Yogesh Kumar[a], Sandeep Kumar Yadav[b], Bibhas Adhikari[c], Ashok Kumar[d]

[a] *Department of EE, Indian Institute of Technology Jodhpur, Residency Road, Ratanada, Jodhpur 342001, Rajasthan, India*
[b] *Electrical Engineering, Indian Institute of Technology Jodhpur, Residency Road, Ratanada, Jodhpur 342001, Rajasthan, India*
[c] *Department of Mathematics, Indian Institute of Technology Kharagpur, India*
[d] *DRDO Delhi, India*

A R T I C L E I N F O

Article history:
Received 28 December 2016
Revised 19 June 2017
Accepted 29 July 2017
Available online 5 August 2017

Keywords:
K-means clustering
Linear regression
OPTICS
Blind signal

A B S T R A C T

Automatic recognition of digital modulation schemes is becoming an active research area in many covert operations. It has many military applications where surveillance and electronic warfare requires a type of modulation in intercepted signal to prepare jamming signals. Most of the approaches are based on modulated signal's component, but the modulation type can be best identified with the use of constellation diagram. The proposed technique is able to recognize M-QAM, M-ASK, and M-PSK modulation scheme in Additive White Gaussian Noise (AWGN) environment. As the constellation points form clusters in the I-Q plane, the order of the modulation can be obtained by estimating the correct number of clusters, which is calculated by OPTICS algorithm. The least square error has been calculated using linear regression from the obtained constellation points, to identify either ASK or PSK and QAM. The error is least for ASK which differentiates ASK from PSK and QAM. To identify between the PSK and QAM, k-means clustering is employed to find the number of centroids equal to order of modulation estimated by OPTICS. With the difference in maximum and minimum absolute value of the centroids, PSK or QAM is recognized. The proposed method shows an improvement in the classification accuracy which reaches 100% using 1024 symbols at 20 dB compared to 98.89%, 98.05%, and 98% when using more complex classifiers like Support Vector Machine, Naive Bayes Classifier, KNN respectively. The method used is unsupervised whereas most of the methods in the literature require training phase to set the thresholds or weights for final model to detect modulation type. This algorithm is also implemented in LabVIEW, and tested on real-time signals. An intelligent system is made which does not require any knowledge of symbol rate, carrier frequency, and any training phase to set thresholds, and detects the type of modulation blindly in real time. Modulated RF signals are generated by NI PXIe-5673 (RF transmitter). NI PXI 5600 is used to downconvert RF signal and NI PXI-5142 (100 MS/s OSP digitizer) is used to sample the downverted signal.

1. Introduction

Automatic modulation classification is the process to classify the modulation format of a signal corrupted by noise and fading channel. It is an intermediate step between signal detection and data demodulation and has significant roles in many fields of military, civil and security applications where friendly signals should be securely transmitted and received, whereas hostile signals must be identified and jammed (Dobre, Abdi, Bar-Ness, & Su, 2007). It is used in link adaptation where modulation used at transmitter adapts according to channel in order to increase transmission reliability. If AMC system is deployed at receiver, it can significantly increase the data throughput and spectrum efficiency by saving bits in each frame containing information of transmitted modulation type. Modulation Identification carries paramount importance in universal Modems and is a precursor to FEC decoding etc. Therefore robust blind modulation identification acts as a foundation of any Communications Intelligence (COMINT) system. AMC is

* Corresponding author. URL: http://home.iitj.ac.in/~sy (Sandeep Kumar Yadav), http://www.facweb.iitkgp.ernet.in/~bibhas/ (Bibhas Adhikari).

E-mail addresses: jajoo.1@iitj.ac.in (G. Jajoo), kumar.33@iitj.ac.in (Y. Kumar), sy@iitj.ac.in (S.K. Yadav), bibhas@maths.iitkgp.ernet.in (B. Adhikari), ak.drdo@gmail.com (A. Kumar).

http://dx.doi.org/10.1016/j.eswa.2017.07.053
0957-4174/

currently an important research area in the design of intelligent radio systems, including Cognitive Radio and Software Defined Radio (SDR).

AMC involves two steps: First, preprocessing of the received signal in which different parameters are estimated like carrier frequency, symbol rate, channel state information, timing and waveform recovery, etc. and the second step is the algorithm to classify the modulation format (Dobre et al., 2007; Hameed, Dobre, & Popescu, 2009). For the latter step, algorithm is broadly grouped into two categories: Likelihood-Based Classifiers and Feature Based methods. Likelihood-Based (LB) approach requires perfect knowledge of channel to give the optimal performance and has high computational complexity compared to feature based approach. In LB method, the likelihood function is calculated for every modulation scheme under consideration and compared to conclude the type of modulation. Maximum Likelihood Estimation (MLE) calculates the likelihood values for modulation hypotheses with all the parameters of the signal known but the modulation format, and classifies the signal to the modulation hypothesis with the maximum value of likelihood function. But how-ever, MLE cannot handle any unknown parameter and to overcome this limitation, Average Likelihood Ratio Test (ALRT) was first used by Polydoros and Kim, 1990 in which the unknown parameter is replaced by its integral in the likelihood function for all of its values. ALRT gets complex when unknown parameters are introduced and also if the model considered is not accurate, the classification becomes suboptimal. Hence Generalized Likelihood Ratio Test (GLRT) is used in which unknown parameters are replaced by Maximum Likelihood Estimates i.e. likelihood function is maximized with respect to unknown parameters, which is computationally less complex. In general, GLRT is a biased classifier in case of nested modulation type like QPSK, 8-PSK, 16-QAM, 64-QAM, thus, Hybrid Likelihood Ratio Test is introduced in which some parameters are handled by ALRT and remaining by GLRT (Panagiotou, Anastasopoulos, & Polydoros, 2000).

Feature Based Methods provide near optimal performance with reduced computational complexity (Dobre et al., 2007). Feature Based method is divided into two steps; first is the feature extraction, in which key features are extracted and the second is pattern recognizer, which identifies modulation type based on the features extracted. Various spectral based features are used by Nandi and Azzouz to classify digital modulation schemes. The features exploit the unique characteristics of different signal modulations in three key signal aspects, namely the amplitude, phase, and frequency. Wavelet Transform based features identify the transients at the phase change during symbol change and also give the time and frequency information, based on which modulation scheme is identified (Ho, Prokopiw, & Chan, 2000; Hong & Ho, 1999). Higher order statistics (cumulants and moments) based features are efficient tools for detection of spectrally equivalent modulations like MPSK and MQAM whose mean and second order statistics are same (Dobre, Bar-Ness, & Su, 2003; Marchand, Lacoume, & Le Martret, 1998; Wu, Saquib, & Yun, 2008). Fourth order cumulants has been used by Swami and Saddler to classify MPSK, MPAM, and MQAM while higher order (up to eighth order) is used in Swami and Sadler, (2000); Wu et al. (2008).

In the feature-based modulation detection, decision making is based on a decision tree. At each level, different features are used and threshold values are optimized which is not very efficient. To overcome such problems, Machine Learning techniques have been employed to identify the modulation type. These are easier to implement and computationally efficient. The feature extraction subsystem calculates the prominent characteristics from the raw data. These features are given to pattern recognizer system which uses machine learning tools viz. KNN classifier, Support Vector Machine (SVM), Neural Network, etc. to classify the signal. Nandi and Azzouz have used Artificial Neural Network along with spectral based features and further Wong and Nandi used ANN with Genetic Programming for modulation classification (Wong & Nandi, 2001). Neural Network does not use the decision tree, instead, it uses nonlinear mapping of features to modulation type. Genetic Algorithm (GA) has been used along with ANN to reduce the dimensionality of feature space (Aslam, Zhu, & Nandi, 2012). Cyclostationary property of linear modulated signal is another feature to detect the modulation type (Dobre & Hameed, 2006). These features along with ANN, LDA, SVM, and KNN are used to classify different modulation schemes (Aslam, Zhu, & Nandi, 2010) and their performances are compared in Satija, Manikandan, and Ramkumar (2014). Overview of feature based classification methods have been given in Hazza, Shoaib, Alshebeili, and Fahad (2013). Other classical approaches, new trends, and comparison of all techniques are given in Dobre et al. (2007).

AMC becomes a challenging task particularly in a noncooperative environment when no prior information of the signal is available. Most of the approaches require channel information or SNR information like in Abdelmutalab, Assaleh, and El-Tarhuni (2016) to set thresholds else their probability of correct classification drops. Also, most of the approaches are not tested practically on real time signal and consider specific order modulation schemes. The proposed approach solves all these problems and has the following advantages:

- It provides higher accuracy of classification compared to most of the approaches in the literature.
- It classifies the modulation scheme without any prior information of SNR. It does not need any training phase to set thresholds.
- The proposed algorithm is also implemented on hardware and tested on real-time signals.
- Method is generalized for any order of modulation in the pool of PSK, QAM, and ASK.

In this paper, Order of modulation has been calculated by Ordering Points To Investigate the Clustering Structure (OPTICS) clustering algorithm, and results show that the proposed algorithm works with 100% accuracy above 10 dB SNR for lower order modulation schemes considered when there is no symbol timing error in the signal received through Additive White Gaussian Noise (AWGN) channel.

The paper is organized in 6 sections. In Section 2, Signal Model and System Model considered in the paper have been explained. In Section 3, the algorithm proposed for the modulation recognition using cluster analyses is explained. Implementation of the algorithm on real time signal is explained in Section 4 followed by the results and discussion in Section 5. We conclude the paper in Section 6.

2. Signal model and system model

It is assumed that the signal is sampled without timing error from rectangular pulses (in simulation). Modulation recognition is a step between the detection of low-level energy signal and full demodulation. In a noisy channel, even if there is a phase error, constellation shape remains unchanged and algorithm is able to classify the modulation type on this basis. Fig. 1 shows the system model to extract the constellation signature from a received signal (Haykin, 2008; Mobasseri, 2000).

Constellation is deformed due to channel noise, carrier frequency, and symbol time estimation error. Channel considered in the present paper is AWGN. Probability distribution function of the Gaussian noise is given by

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad (1)$$