

Studies on Modulation Classification in Cognitive Radios using Machine Learning



Xu Zhu

Department of Communication Engineering and Informatics
The University of Electro-Communications

A thesis submitted for the degree of
Doctor of Philosophy in Engineering

2017 September

Chairperson: Professor Takeo Fujii

1. Member: Professor Xi Zhang

2. Member: Professor Yasushi Yamao

3. Member: Professor Koji Ishibashi

4. Member: Professor Toshiharu Kojima

Day of the Pre-defense: 8th May, 2017

Day of the Final defense: 31st July, 2017

© Copyright by Xu Zhu

Defense Date: 31st July, 2017

All Rights Reserved

概要

周波数資源の逼迫により現在の無線周波数の割り当てでは、将来の無線通信の需要を満足できないことが懸念されており、その解決策として、無線環境の認識結果に従って、送信電力、利用周波数、変調方式などの通信パラメータを適応的に利用することで、周波数利用効率の改善を目指すコグニティブ無線技術に注目が集まっている。コグニティブ無線では、スペクトラムセンシングによりプライマリシステムを検出し、周辺の無線環境を把握することで複数システムが相互干渉なく周波数を共用することが期待されている。このような高度な周波数共用システムの実現には、環境のセンシングや送信に高度な戦略が必要であり、二次利用ユーザにおける信号処理に対する期待が大きくなっている。中でも、スペクトラムセンシングの性能改善は全体の周波数利用効率改善に大きく寄与する。その改善法の一つとして、変調方式の識別を行った上で、プライマリシステムとセカンダリシステムの判定を行う手法がある。例えば、一次システムと二次システムが混在する環境では、二次システムが一次システムの変調を認識することで、干渉の正確な推定が可能となる。また、一次システムの変調の選択状況に応じて、二次システムの送信電力の調整が可能となるなどその活用法は多様である。加えて、信号と雑音の識別も変調方式識別で実現可能であるため、センシングが不可能な環境において信号検出することも可能である。

そこで、本論文では、変調方式識別に注目し、機械学習を活用することでその性能を向上させる手法について検討を行う。1つ目の研究として、教師あり学習による変調方式識別手法の検討を第4章で述べる。ここでは、特徴量として雑音に耐性のある高次キュムラントを用いることとし、ニューラルネットワークを拡張した Stacked Denoising Autoencoder を識別手法として選択する。層構造を持つ事前学習は局所解への収束を回避し、Denoising 機能は雑音耐性を高めることが可能となる。提案手法の有効性を確認するため識別性能と実行速度に関して従来手法と比較をしている。次に、2つ目の研究として、教師なし機械学習を用いた変調方式識別アルゴリズムの検討を第5章で述べる。ここでは、時間周波数分布の特徴量を活用することとし、事前にクラスタ数が決定可能な、Density-based spatial clustering of applications with noise (DBSCAN)を識別手法として利用する。シミュレーションにより、本手法が従来手法より、正確な識別ができることを確認している。加えて、本手法が事前学習を必要としないことから、教師あり学習に比べて汎用性が高いことを示す。最後に第6章で提案する2つの方式を比較し、利点欠点を明らかにする。

今後の課題として、計算ハードウェアの制約の基での、アルゴリズムの最適化があり、その発展策として GPU 計算機を利用した教師あり学習の活用がある。また、あらかじめ準備した変調方式の種類に応じて、ネットワーク構造を再設計する方法についても検討が望まれる。また、教師なし学習については、キャリア周波数オフセットにいかに対応するかが、課題となっている。

Abstract

The current spectrum allocation cannot satisfy the demand for future wireless communications, which prompts extensive studies in search of feasible solutions for the spectrum scarcity. The burden in terms of the spectral efficiency on the radio frequency terminal is intended to be small by cognitive radio (CR) systems that prefer low power transmission, changeable carrier frequencies, and diverse modulation schemes. However, the recent surge in the application of the CR has been accompanied by an indispensable component: the spectrum sensing, to avoid interference towards the primary user. This requirement leads to a complex strategy for sensing and transmission and an increased demand for signal processing at the secondary user. However, the performance of the spectrum sensing can be extended by a robust modulation classification (MC) scheme to distinguish between a primary user and a secondary user along with the interference identification. For instance, the underlying paradigm that enables a concurrent transmission of the primary and secondary links may need a precise measure of the interference that the secondary users cause to the primary users. An adjustment to the transmission power should be made, if there is a change in the modulation of the primary users, implying a noise floor excess at the primary user location; else, the primary user will be subject to interference and a collision may occur. Alternatively, the interweave paradigm that progresses the spectrum efficiency by reusing the allocated spectrum over a temporary space, requires a classification of the intercepted signal into primary and secondary systems. Moreover, a distinction between noise and interference can be accomplished by modulation classification, if spectrum

sensing is impossible. Therefore, modulation classification has been a fruitful area of study for over three decades.

In this thesis, the modulation classification algorithms using machine learning are investigated while new methods are proposed. Firstly, a supervised machine learning based modulation classification algorithm is proposed. The higher-order cumulants are selected as features, due to its robustness to noise. Stacked denoising autoencoders, which is an extended edition of the neural network, is chosen as the classifier. On one hand stacked pre-train overcomes the shortcoming of local optimization, on the other, denoising function further enhances the anti-noise performance. The performance of this method is compared with the conventional methods in terms of the classification accuracy and execution speed. Secondly, an unsupervised machine learning based modulation classification algorithm is proposed. The features from time-frequency distribution are extracted. Density-based spatial clustering of applications with noise (DBSCAN) is used as the classifier because it is impossible to decide the number of clusters in advance. The simulation reveals that this method has higher classification accuracy than the conventional methods. Moreover, the training phase is unnecessary for this method. Therefore, it has higher workability than supervised method. Finally, the advantages and disadvantages of them are summarized.

For the future work, algorithm optimization is still a challenging task, because the computation capability of hardware is limited. On one hand, for the supervised machine learning, GPU computation is a potential solution for supervised machine learning, to reduce the execution cost. Altering the modulation pool, the network structure has to be redesigned as well. On the other hand, for the unsupervised machine learning, that shifting the symbols to carrier frequency consumes extra computing resources.

Acknowledgements

First of all, I would like to give my special thanks to Professor Takeo Fujii. His motivations and encouragements immensely enlightened me throughout my research. It has been a pleasure to perform research under his supervision. I appreciate his guidance, instructions and continuous support during my study period. I learned from him, not only the academic abilities but the philosophy of life. He is the mentor in my life.

I am deeply indebted to other members of this thesis committee: Professor Xi Zhang, Professor Yasushi Yamao, Professor Koji Ishibashi and Professor Toshiharu Kojima. I would like to thank them for their precious comments to improve my thesis and being part of my judging committee.

I would like to thank the kindly help from all friends and the financial support from Japan government.

Finally, I would like to thank my dear wife, Hang Liu, for being with me whether I am happy or lost. She is my spiritual anchor. I enjoy every meal that she cooked for me and every piece of advice that she gave me. Having her by my side, I feel extremely fortunate.

I also want to thank my parents for their continuous support and encouragement. They have been expecting me all goes well. They are my sources of strength and precious wealth.

Contents

List of Figures	vii
List of Tables	xi
1 Introduction	1
1.1 Background	1
1.2 Motivation and Problem	5
1.3 Contributions and Novelty	9
1.4 Scope	11
1.5 Organization of the Thesis	11
2 Signal Model	13
2.1 Signal Models	13
2.1.1 Modulated Signal	13
2.1.2 Signal Model in AWGN Channel	15
2.1.3 Signal Model in Fading Channel	16
2.2 Problem Statement and System Assumptions	17
2.2.1 Problem Statement	17
2.2.2 System Assumptions	18
2.2.3 System Model	18
2.3 Chapter Summary	19
3 Overview of Modulation Classification in Cognitive Radios	21
3.1 General Classes of Modulation Classification	22
3.2 Likelihood-based Methods	22
3.2.1 Maximum Likelihood Method	23

CONTENTS

3.3	Feature-based Methods	25
3.3.1	Time-domain Features and Frequency-domain Features . .	26
3.3.2	High-order Statistics-based Features	29
3.3.3	Cyclostationary Analysis-based Features	31
3.3.4	Classifier	32
3.4	A comparison with Independent Components Analysis (ICA) . . .	36
3.5	Chapter Summary	36
4	Supervised Machine Learning based Modulation Classification	39
4.1	Introduction	39
4.2	Signal Preprocessing and Input Data Constructions	40
4.2.1	Rapid Classification Scenario	41
4.2.2	High Accuracy Classification Scenario	41
4.3	Stacked Denoising Autoencoder	44
4.3.1	SDAE Structure	44
4.3.2	SDAE Forward Propagation	45
4.3.3	SDAE Fine-tuning	47
4.4	Simulation and Results	48
4.4.1	Rapid Classification Scenario	48
4.4.2	High Accuracy Classification Scenario	56
4.5	Chapter Summary	60
4.5.1	Rapid Classification Scenario	60
4.5.2	High Accuracy Classification Scenario	61
4.5.3	Discussion	61
5	Unsupervised Machine Learning based Modulation Classification	63
5.1	Introductions	63
5.2	Feature Extraction	64
5.2.1	Feature-Extraction Tool	64
5.2.2	Amplitude-Feature Extraction	67
5.2.3	Phase-Feature Extraction	68
5.3	Classifier	69
5.3.1	DBSCAN	70

CONTENTS

5.3.2	k -distance plot	71
5.4	Classification Scheme	72
5.5	Simulation and Results	75
5.6	Chapter Summary	81
6	Conclusion	85
6.1	Methods Comparisons	85
6.2	Contribution and Advantages	86
6.2.1	Supervised Classification	87
6.2.2	Unsupervised Classification	88
6.3	Future Research Work	89
6.3.1	Methods Improvement	89
6.3.2	Function Extensions	89
	References	91
	Publications	97

CONTENTS

List of Figures

1.1	Opportunistic spectrum access model for cognitive radio.	3
1.2	Concurrent spectrum access model for cognitive radio.	4
1.3	Main functions of the PHY, MAC and network layers for cognitive radio system.	5
1.4	Implementation of spectrum sensing.	6
1.5	Relationship among modulation classification, energy detection and demodulation of the signal.	7
1.6	Implementation flow of modulation classification based on FB. . .	8
2.1	The constellations of 4QAM in AWGN.	16
2.2	The constellations of 4QAM with phase offset.	16
2.3	System model of the modulation classification in cognitive radio systems.	19
3.1	A typical multi-layer network of ANN.	33
3.2	A typical two-class classification with SVM.	34
3.3	A typical block diagram of MIMO modulation classification. . . .	37
4.1	Data structures of the two scenarios.	43
4.2	The workflow of this method.	43
4.3	Structure of the SDAE with two hidden layers.	46
4.4	Change in the training error corresponding to the different numbers of nodes in each layer.	50
4.5	Classification accuracy comparison of the proposed method with ML.	51
4.6	Classification accuracy for different individual modulations.	51

LIST OF FIGURES

4.7	Influence of the signal sampling synchronization on classification ability.	52
4.8	An ISI channel with a parameter of $[1 \ 0.8 \ 0.3]$ is considered. . . .	52
4.9	Classification accuracy for different individual modulations under ISI.	53
4.10	An ISI channel with a parameter of $[1 \ 0.1]$ is considered.	53
4.11	Classification accuracy for different individual modulations under ISI.	54
4.12	Change in the training error corresponding to the different numbers of nodes in each layer.	56
4.13	Classification accuracy comparison of our method with ML and KNN.	57
4.14	Classification accuracy for different individual modulations.	58
4.15	Influence of the signal sampling synchronization on classification ability.	58
4.16	Influence of the ISI on classification ability.	59
5.1	A signal that contains two symbols in time domain, its contour and the 3D vision on time-frequency domain (PWVD).	65
5.2	Four-distance plot of amplitude extraction and phase extraction. .	70
5.3	Workflow of the signal preprocessing.	72
5.4	An example of amplitude-feature clustering ($\text{SNR} = 10 \text{ dB}$). . . .	74
5.5	An example of phase-feature clustering ($\text{SNR} = 10 \text{ dB}$).	74
5.6	Classification workflow.	74
5.7	Phase-cluster number within <i>MPSK</i> for different SNR.	76
5.8	Amplitude-cluster number within <i>MQAM</i> for different SNR. . . .	77
5.9	Comparison of classification accuracy for our method with ML and KNN.	77
5.10	Influence of the signal-sampling synchronization on individual modulations.	78
5.11	Influence of the phase offset on individual modulations.	78
5.12	Comparison of classification accuracy of our method with ML and KNN against phase offset.	79

LIST OF FIGURES

5.13 Plot of sum-of-square error vs. SNR.	79
5.14 Influence of the roll-off factors on 8PSK.	80
5.15 Influence of the roll-off factors on 64QAM.	80
5.16 Influence of the roll-off factors on classification ability.	81

LIST OF FIGURES

List of Tables

4.1	Higher-order cumulants of the <i>MPSK</i> and <i>MQAM</i>	44
4.2	Parameters used for SDAE (training and classification)	49
4.3	Execution time comparison of proposed method with ML (10,000 iterations)	50
4.4	Execution time comparison of proposed method with ML and KNN (10,000 iterations)	56
4.5	Summary of the three methods	60
5.1	Theoretical numbers of clusters.	73
5.2	Simulation parameters	75
5.3	A summary of the three methods	81
6.1	A comparison between supervised classification and unsupervised classification	86

LIST OF TABLES

Acronym

ASK	Amplitude-shift Keying
FSK	Frequency-shift Keying
PSK	Phase-shift Keying
BPSK	Binary Phase-shift Keying
QPSK	Quadrature Phase-shift Keying
QAM	Quadrature Amplitude Modulation
MPSK	M -ary Phase-shift Keying
MQAM	M -order Quadrature Amplitude Modulation
OFDM	Orthogonal Frequency Division Multiplexing
DSB	Double-sideband Modulation
VSB	Vestigial Sideband Modulation
USB	Upper Sideband Modulation
LSB	Lower Sideband Modulation
SNR	Signal-to-noise Ratio
MC	Modulation Classification
GPU	Graphics Processing Unit
SSE	Sum Squared Error
ML	Maximum Likelihood
FB	Feature-based
KNN	K -nearest Neighbor algorithm
SVM	Support Vector Machine

ANN	Artificial Neural Networks
PWVD	pseudo Wigner Ville Distribution
DBSCAN	Density-based Spatial Clustering of Applications with Noise
SDAE	Stacked Denoising Autoencoders
MHD	Margenau-Hill Distribution
DFT	Discrete Fourier Transform
SCF	Spectral Correlation Function
TFD	Time Frequency Distribution
TFR	Time Frequency Representation
STFT	Short-time Fourier Transform
PSD	Power Spectral Density
PDF	Probability Density Function
PU	Primary User
SU	Secondary User
CR	Cognitive Radio
MAC	Media Access Control
PHY	Physical Layer
LB	Likelihood-based
AWGN	Additive White Gaussian Noise
MIMO	Multi-input Multi-output
ICA	Independent Components Analysis
ISI	Intersymbol Interference

Symbols

S_I, S_Q	in-phase and quadrature components
M	modulation order
f_c	carrier frequency
$W_x(t, f)$	PWVD of a signal
W_{x_1, x_2}	cross-PWVD of x_1 and x_2
W_i^{auto}	auto-PWVD of x_i
$W_0(t, f)$	auto-PWVD of $g(t)$
W_{kl}^I	interference term of x_k and x_l
$g(t)$	pulse shape
x_i	symbol after pulse shape
A_i	modulated symbol
T_s	symbol period
P_{cc}	probability of correct classification
$h(t)$	smoothing function
$s(t)$	transmitted signal
$\omega(t)$	noise
$r(t)$	received signal
$f_\omega(x)$	the probability density function (PDF) of noise

Σ	covariance of noise
σ_I, σ_Q	noise variance of in-phase component and quadrature component
α	factor of amplitude attenuation
θ_0	phase offset
N	number of possible modulations
$s[n]$	modulated sample
$L(s[n])/M, \sigma)$	likelihood function
I_0	modified Bessel function of order zero
N	length of received signal
A_{CN}	centered instantaneous and normalized amplitude of the signal
μ_A	mean value of instantaneous amplitude of one piece of the signal
σ_{ap}	standard deviation of the absolute value of the instantaneous phase
σ_{dp}	standard deviation of the direct instantaneous phase non-linear component
P	the evaluation of the spectrum symmetry
σ_{aa}	standard deviation of the absolute value of normalized and centered instantaneous amplitude
σ_{af}	standard deviation of the absolute value of the normalized and centered instantaneous frequency
σ_a	standard deviation of the normalized and centered instantaneous amplitude
μ_{42}^a	kurtosis of the normalized and centered instantaneous amplitude
μ_{42}^f	kurtosis of the normalized and centered instantaneous frequency
γ_{max}	maximum value of the spectral power density of the centered instantaneous and normalized amplitude

\mathcal{C}_{ab}	cumulant
R_x^α	cyclic autocorrelation
S_x^α	cyclic spectrum (spectral correlation function)
W	weight parameter of each node
b	bias parameters
$\mathcal{C}(x)$	network inputs
j	index of layer
F	activation function
J	average error term of the overall cost function

Chapter 1

Introduction

In this chapter, an introduction to the research on modulation scheme recognition in cognitive radio systems is presented. This chapter is organized as follow. Firstly, the background of the study and related applications in cognitive radio system technologies are reviewed in Section 1.1. Then, in Section 1.2, the motivation and the problems of applying the modulation classification in a thus environment are explained. Section 1.3 describes the contributions and novelties of this thesis. Scope and objectives of the study are shown in Section 1.4. Finally, the organization of the thesis along with the overview of each chapter is given in Section 1.5.

1.1 Background

Spectrum is a spontaneously limited resource for the conventional radio spectrum in the wireless communication systems. In order to support plentiful and various applications and services in a noninterference environment, a fixed spectrum access policy is used by spectrum regulators, which allows them to allocate each bandwidth of current spectrum to one or even more dedicated users. According to the stipulation of the policy, theoretically only the assigned (licensed) users, as known as primary users (PUs) have the right to exploit and utilize the spectrum, while other users, as known as secondary users (SUs) have no permission to use it, even though the PU is not using the spectrum. In the wake of the development of wireless communication services in recent decades, for multiple countries

1. INTRODUCTION

and areas, the majority of the spectrum has been fully assigned, resulting in spectrum scarcity problem. Furthermore, recent researches about the actual utilization of allocated spectrum have revealed that, in most of the time, allocated spectrum is underutilized [1, 2, 3]. Moreover, these researches provide evidence that shows this inflexible and inefficient allocation policy would intensely bring in the scarcity, sometimes, even worse than the basic physical insufficiency of the spectrum. To continue the expansion of the wireless communication industry, a revolution enhancement is necessary to further exploit spectrum resource.

Dynamic spectrum access, therefore, is proposed to meet such challenge as a selective policy to improve the utilization of the allocated spectrum [4, 5].

For the dynamic spectrum access, the available spectrum is still assigned to one or more PUs as in fixed spectrum access. However, these pieces of the spectrum are no longer exclusively dedicated to PUs. Although PUs have the priority of the spectrum utilization, SUs could access to this spectrum temporally whenever PUs are absent or even share this frequency bands with PUs under the controllable interference. Restricted by the dynamic spectrum access policy, PUs could apply the recycle of the radio spectrum under opportunism or share them continually, which could severely contribute to the improvement of the spectrum efficiency.

To support such advanced policy, SUs have to be equipped with the capability of sensing and cognizing the spectrum environment, and such SUs devices are as known as cognitive radio (CR) [4, 5] or cognitive radio users. Various CRs devices with various spectrum sensing capacity are built for different kinds of purposes. For instance, a cognitive radio device should sense whether the signal transmitted from PU exists or not [6], or even be able to calculate its own interference level towards PU at the receiver (Rx) end. For some particular situations, intelligence CR devices are allowed to obtain some unclassified information about the PU from the PU transmitter (Tx). However, it may cost CR users great calculation and processing time for implementing, sensing, and cooperation etc.

According to different objects, CR users could utilize or access the allocated spectrum in two basic schemes of cognitive spectrum access models. One is called the opportunistic spectrum access model, as we can see in Figure 1.1, which realizes the spectrum sensing by detecting the holes in the spectrum. It

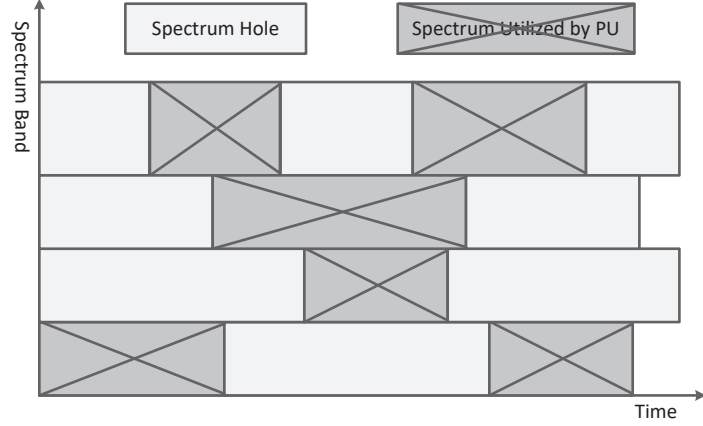


Figure 1.1: Opportunistic spectrum access model for cognitive radio.

reveals that SUs can use the spectrum hole if PUs are absent. Meanwhile, SUs may be required to reset their own carrier frequency or modulation to fit such transmission parameters. In order to realize the implementation, the CR uses need to supervise related spectrum frequently and be capable to evacuate the spectrum as soon as PU reappears. The other CR implementation policy is called the concurrent spectrum access model, as we can see in Figure 1.2. From the figure, one can notice that the CR users are existing while PUs are staying and applying their licensed spectrum. The CR users are demanded to control their interference power of the transmitting towards PU Rx and ascertain the interference being less than the thresholds which are the limitation of interference that PUs can tolerate. Therefore, the CRs applying concurrent spectrum access model needs to have the capability of calculating the interference power at some specific space point.

In an actually cognitive radio system, multiple CR users coexist, which makes it similar to the general wireless communication system. Moreover, the cognitive radio networks stay in two classes that including basic infrastructure network and ad hoc network [7]. A cognitive radio network system can be described as an intelligent network of various coexisting or non-coexisting single layer network of CR user overlaying together. Establishing a fully functional cognitive radio

1. INTRODUCTION

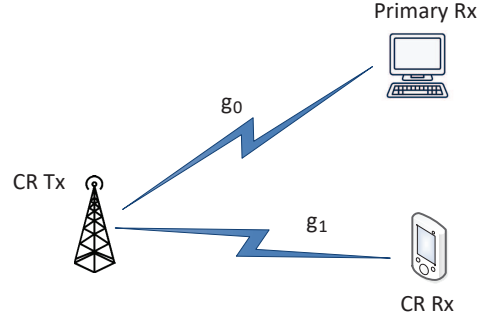


Figure 1.2: Concurrent spectrum access model for cognitive radio.

network could be a very challenge work, on account of melting all the various network of cognitive radio users. Because it is necessary to fulfill the demand for the signal and algorithm processing of physical (PHY) layer, spectrum management and controlling of medium access control (MAC) layer as well as network layer routing and statistical control. Even more, all the managing, calculating and controlling systems cause the complicated influence on each other, which needs more care to design cross layer or frameworks of controlling.

As we can see in Figure 1.3, it shows the relationship between PHY layer, MAC layer and network layer in a cognitive radio system. For a physical layer, the main task is to realize the spectrum sensing, which needs to detect spectrum holes and collect more information about channel condition and channel gain after sensing the transmission environment. According to the obtained information after spectrum sensing, the cognitive MAC layer is applied to optimize and re-constitute the network between the transmitter and receiver, which contains the tasks of spectrum aware access control. A CR user has to be able to look after the tradeoff between the access opportunity and the cognitive sensing requirement. The operator of these two functions is called sensing access coordinator. Regarding the network layer, the main three tasks are the quality of service, network tomography, and spectrum aware routing [8]. Overall, the PHY layer, MAC layer, and network layer are linked together to give an accurate access to the spectrum holes dynamically and efficiently. The model that I described above, however, is not the only model for the cognitive radio network. [7] offers more CR models.

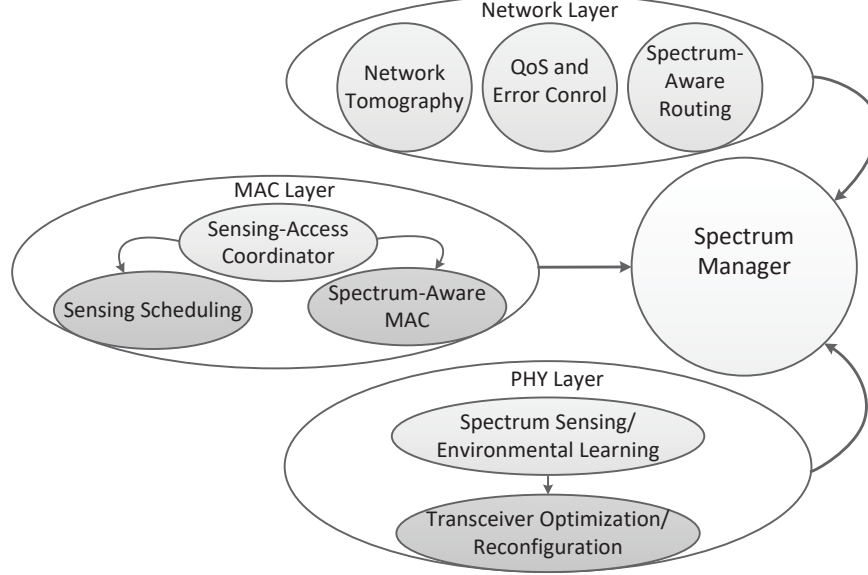


Figure 1.3: Main functions of the PHY, MAC and network layers for cognitive radio system.

1.2 Motivation and Problem

As the description above, spectrum hole means that the spectrum allocated to the primary user is underutilized at this moment or this location. For some particular spectrum bands, such as TV band, the TV programs are broad-casted strictly according to the timetable. It implies that in some regions, the utilization of spectrum band could be predicted by the secondary users with some helpful solutions, i.e. database. If the SUs are unable to reach out such information or the PU utilizes the spectrum prospectively, the spectrum sensing [9, 10, 11, 12] could greatly enable the CR users to recognize the idle spectrum in order to protect the interests of PU, as we can see in Figure 1.4. According to the figure, the primary transmitter is using the assigned spectrum band to communicate with the certain receiver while the CR users need to access to this band for communication. To avoid the interference towards PU, the SUs have to sense the spectrum to find out the idle spectrum. In addition, the idle spectrum can be detected directly after finding out the primary user receiver which is called direct spectrum sensing.

The performance of the spectrum sensing can be extended by a robust mod-

1. INTRODUCTION

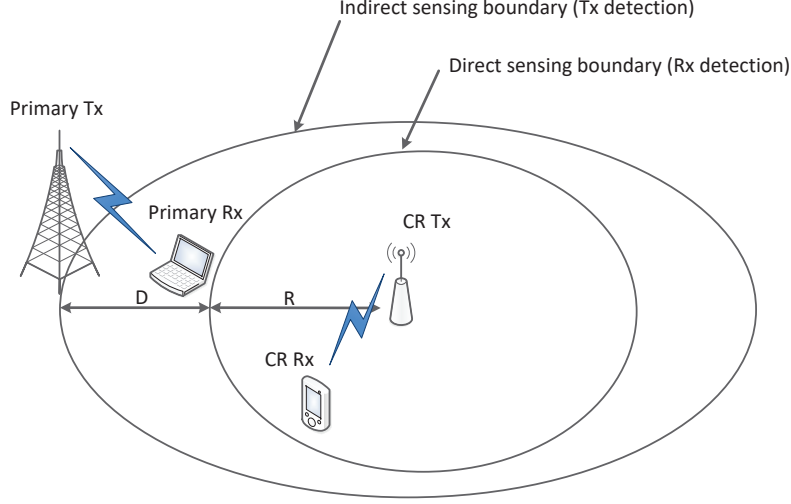


Figure 1.4: Implementation of spectrum sensing.

ulation classification (MC) scheme to distinguish between a primary user and a secondary user along with the interference identification. For instance, the underlying paradigm that enables a concurrent transmission of the primary and secondary links may need a precise measure of the interference that the secondary users cause to the primary users. There should establish a supervision of the transmission power, and if there is a change in the modulation of the primary users, implying a noise floor excess at the primary user location; else, the detection performance of primary user will be severely deviation due to interference which may cause collision. Alternatively, the interweave paradigm that progresses the spectrum recycling by utilizing the allocated spectrum over a temporary space, requires a classification towards the intercepted signal into primary and secondary systems. Moreover, when spectrum sensing fails, modulation classification could accomplish classification between the noise and the interference. Therefore, there is a recent surge in the application of CR accompanied by the MC. As a flexible implementation of a specific task is used by the CR, the modulation scheme that appeared less flexible becomes a resourceful process. This requirement calls for modulation classification as an intermediate step before the demodulation in CR networks [13]. Modulation classification has been a fruit-

ful area of study for over three decades [14] is due to its applicability to several practical problems. The general relationship among the signal classification, the signal detection, and the common demodulation is given shown in Figure 1.5. The classification is the intermediate steps between detection and demodulation. The more information one needs for the processing, the more they will give after the processing. For instance, only the approximate frequency and spectrum band are required for the energy detection, but the exact center frequency, spectrum band, utilized modulation and the parameters of modulation are required for demodulation. Therefore, the modulation classification could be moderate for the received signal processing.

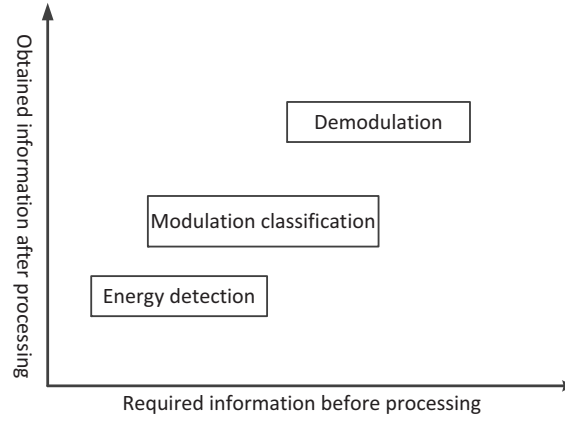


Figure 1.5: Relationship among modulation classification, energy detection and demodulation of the signal.

The classifiers of the MC in literature can be broadly categorized into two categories [15]: likelihood-based (LB) method and feature-based (FB) methods. The LB transforms the MC into a hypothesis testing and compares the likelihood probability with a threshold. In order to obtain the likelihood probability, all channel parameters have to be known. It can be seen from most likelihood-based papers that the LB methods require an estimation of parameters [16, 17, 18, 19] such as for e.g., the form of the distribution function and mean value, the variance as well as the signal-to-noise ratio (SNR) level. The advantage of the LB method is not only to guarantee the best classification results under the Bayes minimum miscalculation cost criteria theoretically but also able to obtain the classification

1. INTRODUCTION

performance curve through theoretical analysis. The LB method classification performance would be the upper limit in theory which could be used to testify the performance of the modulation classification of feature-based methods. However, due to the requirement of the prior information, it leads to considerable difficulty in determining the decision threshold, often with an ideal probability of false alarm. Moreover, as the cost of such an estimation is prohibitive, it makes the finding of an optimal solution using an exhaustive search infeasible.

Feature-based methods, conversely, doesn't need an estimation of parameters [20, 21]. Therefore, they are usually less complicated and easy to utilize, in spite they may be sub-optimal. However, an optimal performance can be achieved with a proper design. As we can see in Figure 1.6, it shows a flow path of an FB classification implementation flow. To provide the assembly of the classification models of demand, first, one has to select the features of classification and the rules of classifiers, then extract the parameter information from the signal samples of known modulation types. The information of feature is used to train the classifier until the output of the classifier satisfies the error requirement, or directly set the threshold and classification function of the classifier based on the statistical analysis, i.e. mean value and variance. For the testing phase, the classification performance is based on the successful classification probability using the feature extracted from another set of communication signal samples. The FB can be generally decomposed into two moves: feature extraction that is normally implemented in the time-domain or frequency domain and a classifier that can be supervised or unsupervised [22].

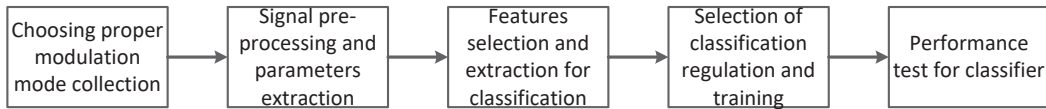


Figure 1.6: Implementation flow of modulation classification based on FB.

1.3 Contributions and Novelty

The contributions and novelties of the thesis are considered from two main aspects. The modulation classification approaches we used here includes the unsupervised processing and the supervised training.

For the supervised approach, the contributions and novelties are as follow:

- A rapid classification scenario is considered here. So far, there is few of paper that mentions this scenario, although the processing time is the priority and classification accuracy can be compromised in special conditions. Performance evaluation shows a significant speed advantage over the conventional maximum likelihood (ML) method.
- In a rapid classification scenario, expert features are not necessary. Therefore, feature extraction that is compulsory in most conventional methods is omitted. It simplifies the procedure of the modulation classification and renders rapid classification more achievable.
- A high accuracy classification scenario is also considered in the thesis. Although expert features are utilized as network inputs, the stacked denoising autoencoder as the classifier improves the noise resistance performance. Performance evaluation shows an accuracy advantage over the conventional ML method and the feature-based method.
- So far, there is few deep networks based literature that exhibits the influence of the signal sampling synchronization over the classification ability, although it is a significant issue in the modulation classification field. The influence of the timing offset over the classification accuracy are investigated in both the scenarios. This ensures a comprehensive evaluation to understand the ability of this method. In the high accuracy scenario, the proposed method is robust to the timing offset.
- Not only a ready network structure but also the selection methods of the network structure are presented.

1. INTRODUCTION

- The performance of the proposed approach is also investigated on different individual modulations. This gives insights into the classification ability of this method.

For the unsupervised method, the contributions and novelties are as follows:

- It has better performance with respect to its noise resistance which is based on the proposed new features using time-frequency distribution. These features show good robustness to noise. The performance evaluation shows an accuracy advantage over the conventional ML method [19] and K-Nearest Neighbors (KNN) method [23]. The novel method is robust to phase offsets, which always degrade the performance of likelihood-based methods [19] and KNN method [23] in the classification of phase shift keying (PSK) and quadrature amplitude modulation (QAM) modulation schemes.
- To date, there are few reports in the literature that exhibit the influence of the signal sampling synchronization over the classification ability, although it is a significant issue in the field of modulation classification. The thesis investigated the influence of the timing offset on the classification accuracy. This ensures a comprehensive evaluation to understand the capabilities of this method. This method shows good robustness to timing offsets for m -ary phase shift keying (MPSK).
- It investigates the clustering validation against SNR and showed that density-based spatial clustering of applications with noise (DBSCAN) is valid for this method.
- DBSCAN does not require the time-consuming training of the classifier. This is a significant advantage over supervised classifiers when rapid processing is expected.
- In addition, it is simple to obtain the parameters of this method. On the contrary, the cumulant-based approach needs a complicated process to determine its decision rules [20].

1.4 Scope

Although, the cognitive radio system contains various aspects of communication technologies, such as the physical layer, the medium access layer, the network layer routing and statistical control. The modulation classification in this thesis is the technology for the physical layer.

In this thesis, as a hierarchical classification system, the modulation classes have to be identified, after which the modulation order can be determined. In a CR system, some fundamental information about the primary user is often accessible. Moreover, because digital modulation has better immunity to noise, binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), 8PSK, 16QAM, as well as 64QAM are widely used in CR systems. Therefore, they are mostly discussed in literature pertaining to modulation classification. Throughout this thesis, I assume that there is a single carrier-transmitted signal, whose possible modulation type includes BPSK, QPSK, 8PSK, 16QAM, and 64QAM, and the goal is to classify the modulation type that it is using. For particular classifier of the deep learning network, the complex symbols rather than pulse shaped complex signals as the network input are proposed, simplifying the network topology and reducing the calculation overhead. This can be easily verified if our network structure is compared with other convolutional neural network-based methods [24] and conventional neural network-based methods.

To date, there is little literature that reports the influence of the signal-sampling synchronization over the classification ability, although it is a significant issue in the modulation classification field. The influence of the timing offset is investigated over the classification accuracy as well as the phase offset.

1.5 Organization of the Thesis

This thesis gives the summarization of the research on the both the supervised and unsupervised modulation classification for a transmission signal from PU received at the CR user received ends. This thesis consists of 6 chapters as follows.

- Chapter 1 introduces the background, the motivation, the objection, the contributions, the novelty, and the scope of the studies.

1. INTRODUCTION

- Chapter 2 explains the signal models and system models of the research. Mainly, this research is concerned about classifications in AWGN channel. However, to better evaluate the research, phase offset, timing offset, and ISI channel are taken into consideration.
- Chapter 3 provides an overview of modulation classifications, including likelihood based methods and features based methods. The purpose of this chapter is to provide a preliminary of the following proposal chapters.
- Chapter 4 presents a supervised machine learning based modulation classification method. Higher order cumulants are used as extracted feature. Stacked denoising autoencoder is used as the classifier. Two scenarios have been considered in this chapter, which are rapid classification scenario and high accuracy classification scenario.
- Chapter 5 presents an unsupervised machine learning based modulation classification method. Time-frequency analysis is used for feature extractions. Density-based spatial clustering of applications with noise is used as the classifier. The proposed method here shows a stronger ability of classification than conventional methods.
- Chapter 6 concludes this thesis, and summarize the contributions of the proposals in chapter 4 and 5. Future work is discussed in this chapter as well.

Chapter 2

Signal Model

In this chapter, the expression form of modulated signal is first discussed, such as QAM and PSK, because this research mainly focuses on classifications of these two modulations. The difference is compared between their expressions to enhance the features that could be utilized for classifications. Signal models are the fundamental link to all kinds of modulation solutions. Most of the likelihood based methods rely on a certain signal model for decision making. Moreover, although some of the feature-based methods doesn't need a signal model hypothesis, a known signal mode is still able to provide assistance in parameter selections for classifiers. Therefore, signal models in additive white Gaussian noise (AWGN) and fading channel were presented.

This chapter is organized as follow. Section 2.1 describes the signal models. Concretely, the expressions of modulated signal are given in section 2.1.1. Then the signal model in AWGN channel and fading channel are discussed. In section 2.2, problem statement and system assumptions are discussed. The system model in section 2.2 is briefly introduced as well.

2.1 Signal Models

2.1.1 Modulated Signal

A general model for the signal and the definitions of the relevant modulations are first considered. Over one symbol interval, T_s , the transmitted signal can be

2. SIGNAL MODEL

written in a quadrature form as follows:

$$s(t) = s_I \cos(2\pi f_c t) - s_Q \sin(2\pi f_c t), \quad (2.1)$$

where s_I and s_Q are the in-phase and quadrature components, respectively, of the baseband.

For an *MPSK*, where M is the order, all the information is carried by the phase of the signal and the minimum phase shift between two adjacent symbols is $2\pi/M$. Considering this, an *MPSK* modulated signal is given by:

$$\text{PSK}(t) = \sum_i g(t - iT_s) \cos(2\pi f_c t + (2m + 1) \frac{\pi}{M}), \quad (2.2)$$

where the initial phase is assumed to be zero with $0 \leq m \leq (M - 1)$.

For the m -ary quadrature amplitude modulation (*MQAM*), the information is carried by both the amplitude and the phase of the signal. In a quadrature form, the following equations can be established for *MQAM* modulated signals:

$$\text{QAM}(t) = s_I \cos(2\pi f_c t) + s_Q \sin(2\pi f_c t), \quad (2.3)$$

$$\begin{aligned} s_I &= \sum_i A_i g(t - iT_s), \\ s_Q &= \sum_i B_i g(t - iT_s), \end{aligned} \quad (2.4)$$

where A_i and B_i take values of $(2i - 1 - L)d$ with $i = 1, 2, \dots, L$. L is the modulation order and d is the constellation distance.

In order to give the *MQAM* a specific contrast with the *MPSK*, it will be beneficial to rewrite (2.3) as:

$$\text{QAM}(t) = \sum_i g(t - iT_s) C_i \cos(2\pi f_c t + \varphi_i), \quad (2.5)$$

where $C_i = \sqrt{A_i^2 + B_i^2}$ and $\varphi_i = \arctan(B_i/A_i)$.

2.1.2 Signal Model in AWGN Channel

In most of the literature, AWGN is considered as noise model for the performance evaluation in modulation classifications. Based on this assumption, noise is the only limitation of accurate classification. In order to evaluate the performance of the proposed methods justly, AWGN is considered as noise model as well. However, to better reveal the performance of this methods, the performance is still presented in the case of timing offset and phase offset at the end of chapter 4 and chapter 5.

A general expression for a signal in AWGN channel is given by:

$$r(t) = s(t) + \omega(t), \quad (2.6)$$

where $s(t)$ and $\omega(t)$ represents signal and noise, respectively.

For complex signal and corresponding complex noise, the probability density function (PDF) of noise is given in (2.7):

$$f_{\omega}(x) = \frac{1}{2\pi\sqrt{\Sigma}} e^{-\frac{|x|^2}{2\sqrt{\Sigma}}}, \quad (2.7)$$

where, Σ is the covariance of noise.

In some literature, in-phase component and quadrature component are considered separately. Therefore, it is necessary to discuss the noise PDF of in-phase component and quadrature component, respectively. The covariance of in-phase component and quadrature component is given by:

$$\Sigma = \begin{bmatrix} \sigma_I^2 & \rho\sigma_I\sigma_Q \\ \rho\sigma_I\sigma_Q & \sigma_Q^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}, \quad (2.8)$$

where, σ_I and σ_Q are variance of in-phase component and quadrature component, respectively. This equation reveals that the variance of in-phase component and quadrature component can be replaced by σ , a identical variance. This is because in-phase component and quadrature component are orthonormal to each other. The projection of noise being independent is obtained.

The PDFs of in-phase component and quadrature component are given by:

$$f_{I(\omega)}(x) = f_{Q(\omega)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{|x|^2}{2\sigma^2}}. \quad (2.9)$$

The constellations of 4QAM in AWGN are given in Figure 2.1, where SNRs are 0dB and 10dB, respectively.

2. SIGNAL MODEL

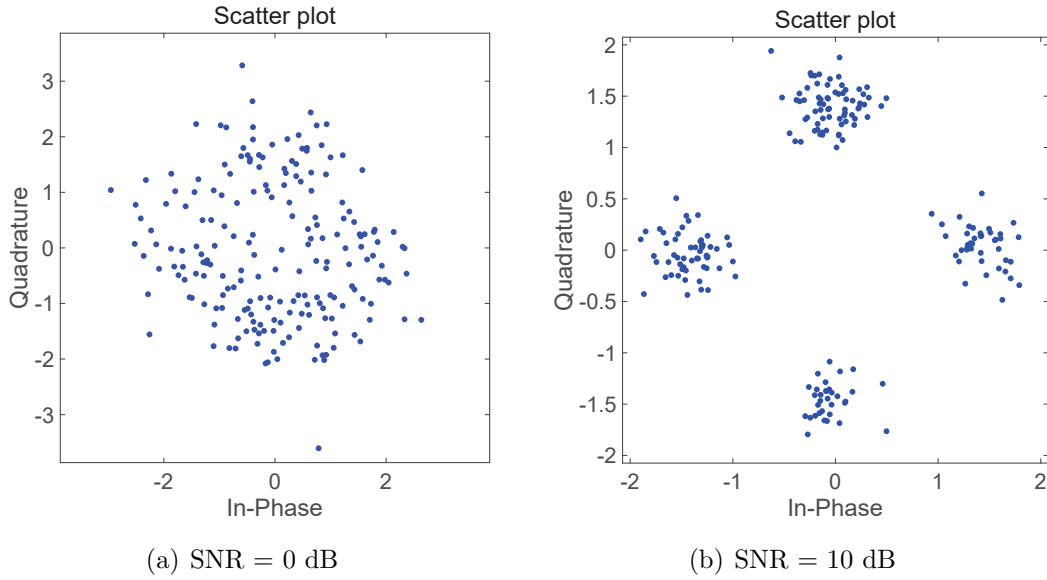


Figure 2.1: The constellations of 4QAM in AWGN.

2.1.3 Signal Model in Fading Channel

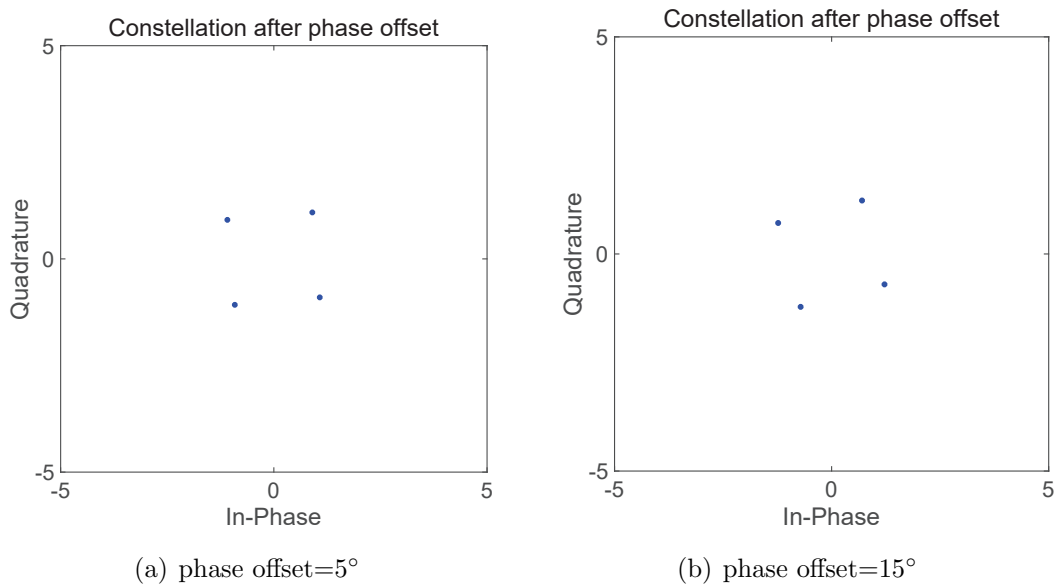


Figure 2.2: The constellations of 4QAM with phase offset.

In wireless communication, the fading channel is an unavoidable issue. More-

2.2 Problem Statement and System Assumptions

over, academics have proposed a variety of fading channel models, in order to quantify the effect of fading. However, till date, there are few methods aiming at the specific fading model. Therefore, instead model-related derivations, this research concentrates on the consequences of fading channel, such as phase offset.

A general expression for a signal in fading channel is given by:

$$r(t) = \alpha e^{j\theta_0} s(t) + \omega(t), \quad (2.10)$$

where, α is the factor of amplitude attenuation, and θ_0 represents phase offset. These two factors are considered as constant since the symbol period value is much smaller than the rate of change in most cases.

The constellations of 4QAM with phase offset are given in Figure 2.2, where phase offset are 5° and 15° , respectively. Apparently, phase offset caused rotation of constellation upon the origin. It can be expected that phase offset affects the performance of some methods, indeed.

2.2 Problem Statement and System Assumptions

Because digital modulation has better immunity to interference [25], it is mostly discussed in the literature regarding modulation classification. Therefore, the research focuses on classifications of digital modulation in this thesis as well.

2.2.1 Problem Statement

In digital modulations, amplitude, phase, and frequency can carry information. Therefore, one may have various modulation methods, such as m -ary amplitude shift keying (*MASK*), *MPSK*, m -ary frequency-shift keying(*MFSK*), and *MQAM*. On one hand, there is an increasing need for the spectrum for IoT devices, which make stringent requirements on spectrum utilizations. Therefore, *MFSK* is not the focus of the discussion. On the other hand, *MASK* is rarely used because it is extremely sensitive to noise. Therefore, *MASK* is not the focus of the discussion, neither.

2. SIGNAL MODEL

Only *MPSK* and *MQAM* are taken into account in this thesis. Here, this research assumes that there is a single carrier-transmitted signal, whose possible modulation types include the *BPSK*, *QPSK*, *8PSK*, *16QAM*, and the *64QAM*. The mission is to classify the signals used by their modulation types.

2.2.2 System Assumptions

Assume that the modulation types are obtained from a set of N possible modulations, where $M = M_1, M_2, \dots, M_N$. Let P_c donate the probability that the classification result is identical to the transmitted signal. Then, using the conditional probability, P_c can be expressed as:

$$P_c^{n|n} = P(D = M_n | M_n), \quad (2.11)$$

where $D = M_n$ represents the case, wherein, the classification result is M_n . Then, the average probability of obtaining a correct classification is given by:

$$P_{cc} = N^{-1} \sum_{n=1}^N P_c^{n|n}. \quad (2.12)$$

P_{cc} is utilized to evaluate the performance of all the classifications throughout the thesis.

2.2.3 System Model

The system model is depicted in Figure 2.3. The secondary user intercepts the signals from the primary user when a primary user is present. After a series of preprocessing steps, the symbols for the subsequent modulation classification can be acquired.

There are several MC methods that assume sampling synchronization, carrier recovery, and waveform recovery has been achieved before MC. However, to better understand the performance of the proposed method, the influence of the signal sampling synchronization is exhibited over the classification ability.

This research emphasizes that although there are papers [26] [24] that utilize deep networks for signal modulations, they use pulse shaped signals rather than

modulated symbols. Therefore, more input nodes are compulsory for their networks, entailing a significant calculation time. This problem is avoided by using modulated symbols in this study.

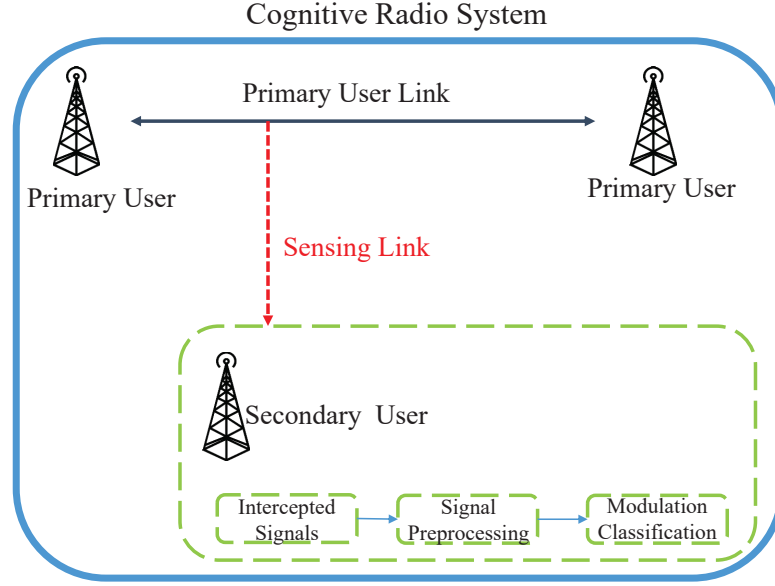


Figure 2.3: System model of the modulation classification in cognitive radio systems.

2.3 Chapter Summary

In this chapter, the signal models is first discussed to provide a chance of a fair comparison with other methods. Signal models in AWGN and fading channel are presented. Then the main problem that is concerned about has been presented. Specifically, this research concentrates on classifications of digital modulations, such as *MPSK* and *MQAM*. This chapter also explains the reason why this research is focusing on these modulations. In addition, the system assumptions and system models are given at the end of this chapter, to better explain potential application scenario. It should be pointed out that although this thesis is about modulation classification in cognitive radio scene, the proposed methods are universal for other scenarios.

2. SIGNAL MODEL

Chapter 3

Overview of Modulation Classification in Cognitive Radios

Modulation classifications are first innovated for military applications, such as for e.g., electronic ware. It can be utilized to jam signals through launching interference with the same modulation mode. Moreover, the classification of signal modulation is required to crack the intercepted signals. With the development of communication technology, the adaptive modulation has been widely used by adaptive transmitters. Specifically, adaptive transmitters can alter the modulation modes according to the channel estimations.

This chapter is an overview to the fundamental principle of modulation classification. Two general classes of modulation classification are involved: likelihood-based and feature-based methods. To better understand the proposed methods, the two steps of the likelihood-based method are firstly introduced: likelihood evaluation and likelihood comparison. Then, the two steps of the feature-based method are further explained: feature extractions and classifiers, since this research mainly concentrate on this class.

This chapter is organized as follow. Section 3.1 introduces the two general classes of modulation classification. Section 3.2 presents the theory of likelihood-based methods. Then Section 3.3 explains the structure of feature-based methods, which contain feature extractions and classifiers. The differences between ICA and machine learning are discussed in Section 3.4.

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

3.1 General Classes of Modulation Classification

The modulation classifications in literature can be broadly categorized into two categories: likelihood-based and feature-based methods.

On one hand, the likelihood-based method transforms the modulation classification into a hypothesis testing and compares the likelihood ratio against a threshold. Therefore, the likelihood-based method is supposed to be the optimal detector in Bayesian classification. However, in order to obtain the likelihood probability, all channel parameters have to be known. It can be seen from most likelihood-based papers that the likelihood-based methods require the estimation of parameters such as for e.g., the SNR level. This leads to considerable difficulty in determining the decision threshold, often with an ideal probability of false alarm. Moreover, as the cost of such an estimation is prohibitive, it makes the finding of an optimal solution using an exhaustive search infeasible.

The feature-based methods, on the other, and utilize the features extracted from raw signals and employ suitable classifiers to make a decision. Although, the feature-based methods are not optimal in most cases, easy to implement has always been one of many advantages, because they don't need the estimation of parameters. In fact, an optimal performance can be achieved with a proper design.

3.2 Likelihood-based Methods

As one of widely used classification methods, likelihood-based methods is famous for its classification accuracy if all required information of channel is known [17].

The likelihood-based method includes two steps: likelihood evaluation and likelihood comparison. In the step of likelihood evaluation, the likelihood is derived for every modulation, according to the samples of intercepted signals. Specifically, a signal model has to be made in advance, then likelihood value can be derived. In general, the computation complexity of this step is nonnegligible, which includes a lot of execution cost. An optimization of a signal model

can reduce the computation complexity, which is a hot spot for likelihood-based methods [17, 25].

In the step of likelihood comparison, the likelihood value of different modulations is compared. Early likelihood-based methods are identical to binary hypotheses, where the likelihood value of different modulations is compared with a ratio. The optimization of the threshold can further improve the classification accuracy, although it is not effortless to select the threshold. An intuitive way of making a decision is to select the hypothesis that has the largest likelihood as the classification result. Apparently, it is an easier solution for implementation. There have been many studies that proposed valid approaches to improve the classification accuracy and simplify the calculation complexity. In this thesis, maximum likelihood method is chosen as a representative of likelihood-based methods. The principle of maximum likelihood method is presented and applied to compare the proposed methods with this method in chapter 4 and chapter 5.

3.2.1 Maximum Likelihood Method

Maximum likelihood method selects the hypothesis that has the largest likelihood as the classification result. All parameters have to be known in advance, including channel knowledge, although this seems to be a less practical assumption. Firstly, this section is started with likelihood function of the AWGN channel.

For a modulated sample $s[n]$, the likelihood function $L(s[n] | M, \sigma)$ corresponding to modulation M can be presented in terms of a conditional probability in equation:

$$L(s[n] | M, \sigma) = p(s[n] | M, \sigma), \quad (3.1)$$

where σ is the noise variance. Moreover, the conditional probability in AWGN channel can be further expressed by the PDF. Therefore, (3.1) can be expressed by:

$$L(s[n] | M, \sigma) = \sum_{i=1}^M \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|s[n]-A_i|^2}{2\sigma^2}}, \quad (3.2)$$

where A_i is the modulated symbol. Regarding a series of intercepted samples, using all intercepted samples, the joint likelihood is calculated by multiplying all

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

individual likelihood of each sample, which is given in:

$$L(s[n] | M, \sigma) = \prod_{n=1}^N \sum_{i=1}^M \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|s[n]-A_i|^2}{2\sigma^2}}, \quad (3.3)$$

here, N is sample length. In most case, the absolute value of the likelihood function is relatively small. Therefore, likelihood functions are generally given in logarithmic form. Then equation (3.3) can be rewritten in terms of logarithmic expression:

$$\begin{aligned} \log L(s[n] | M, \sigma) &= \log \left(\prod_{n=1}^N \sum_{i=1}^M \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|s[n]-A_i|^2}{2\sigma^2}} \right) \\ &= \sum_{n=1}^N \log \left(\sum_{i=1}^M \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|s[n]-A_i|^2}{2\sigma^2}} \right). \end{aligned} \quad (3.4)$$

Specifically, the likelihood is identical to PDF of samples. And the PDF can be derived from different aspects of samples, such as amplitude and phase. Here, PDF of phase is taken as an example. In AWGN channel, the likelihood function of the signal phase can be found in:

$$\begin{aligned} L_{\theta(s)}(s[n] | A_i) &= \frac{e^{-\frac{|A_i|^2}{2\sigma^2}}}{2\pi} + \\ &\frac{|A_i| \cos(\theta(s[n]) - \theta(A_i))}{2\sigma\sqrt{2\pi}} [1 + E] e^{-\frac{|A_i|^2}{2\sigma^2} \sin^2(\theta(s[n]) - \theta(A_i))}, \end{aligned} \quad (3.5)$$

where, E is defined by:

$$E = \operatorname{erf}(|A_i| \cos(\theta(s[n]) - \theta(A_i))). \quad (3.6)$$

Here, erf is the error function. $\theta(A_i)$ is the phase of symbol A_i .

On one hand, signal phase is an intuitive information, which can be implemented within *MPSK* classification and *MQAM* classification [27]. Moreover, signal phase is more robust to noise than amplitude information, which makes it widely used by lots of literature. However, as mentioned above, all parameters and channel information are necessary for likelihood methods, which is hard to fulfill in practice. For instance, phase offset is a common condition which cannot

be ignored. Therefore, likelihood methods are significantly affected by the phase offset.

Due to the high complexity of the likelihood function, a lot of simplified methods are presented. For example, the PDF calculation method using Von Mises distribution is a typical simplified solution, which can be used for phase likelihood in equation:

$$L_{\theta(s)}(s[n] | M, \sigma) = \prod_{n=1}^N \sum_{i=1}^M \frac{1}{M} \frac{e(|A_i|^2 / \sigma^2) \cos(\theta(s[n]) - \theta(A_i))}{2\pi I_0(|A_i|^2 / \sigma^2)}, \quad (3.7)$$

where, I_0 denotes the modified Bessel function of order zero. It must be noted that the PDF calculation method using Von Mises distribution deviates from the real PDF if signal is heavily noised.

On the other hand, the amplitude information can be relied within MQAM classification and multiple pulse amplitude modulation (MPAM) classification [28]. Moreover, the amplitude information is robust to phase offset and carrier offset, which is an important advantage over phase information. This can be easily expected because phase offset only cause the constellation to rotate. the amplitude likelihood can be expressed in equation:

$$L_{|r|}(r) = \prod_{n=1}^N \sum_{i=1}^M \frac{1}{M} \frac{|A_i|}{\sigma^2} e^{-(r^2 + |A_i|^2) / 2\sigma^2} I_0\left(\frac{r |A_i|}{\sigma^2}\right). \quad (3.8)$$

3.3 Feature-based Methods

The FB can be generally decomposed into two moves: feature extraction that is normally implemented in the time-domain or frequency domain and a classifier that can be supervised or unsupervised. After the detailed analyzation and discussion of the property of modulation implementation, its not difficult to find out the crucial features for modulated signal which applying some particular modulation mode.

In this section, part of the easily identified features for MC are listed. In the front parts, the spectrum-based features that develop the spectrum features for various signal components are investigated here. At the same time, the high-order statistics-based features are exploited for modulations of various orders and

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

models of signals. The cyclostationary analysis-based features according to the cyclic spectrum features are also discussed. In addition, some main feature-based classifiers are given at the end of the section.

3.3.1 Time-domain Features and Frequency-domain Features

The classification methods for the basic analog as well as the digital modulations applying some spectrum-based features are presented by Azzouz and Nandi [28] at the end of the 20th century. They summarize and improve the previous research of Fabrizi, Lopes and Lockhart, Chan and Cadbois, etc [29]. They investigate unique features of three main behaviors of the amplitude, phase, and frequency of various modulations. These three behaviors could show different features for each type of modulation, after building a perfectly full collection pool of wanted modulations, the most valid features could help distinguish them. The decision-making tree is utilized to give an implementation flow for classification processing.

The maximum value of the spectral power density of the centered instantaneous and normalized amplitude is expressed as:

$$\gamma_{max} = \frac{\max |DFT(A_{cn})|^2}{N}, \quad (3.9)$$

where, N stands for the number of samples of received signal, A_{cn} stands for the centered instantaneous and normalized amplitude of the signal, and DFT is short for discrete Fourier transform. At this point, the centered instantaneous and normalized amplitude which is denoted to be the compensation of the unknown channel transmission attenuation, is expressed as:

$$A_{cn}[n] = A[n] - 1, \quad (3.10)$$

where, $A_n[n] = A[n]/\mu_A$, and μ_A stands for the mean value of instantaneous amplitude of one piece of the signal, according to the expression of:

$$\mu_A = \frac{1}{N} \sum_{n=1}^N A[n]. \quad (3.11)$$

The standard deviation of the absolute value of the instantaneous phase, namely σ_{ap} , can be expressed by:

$$\sigma_{ap} = \sqrt{\frac{1}{N_c} \left(\sum_{A_n[n] > A_t} \phi_{NL}^2[n] \right) - \left(\frac{1}{N_c} \sum_{A_n[n] > A_t} |\phi_{NL}[n]| \right)^2}, \quad (3.12)$$

where, N_c is the number of the samples of received signal under the condition: $A_n[n] > A_t$. A_t is set to be the threshold that would eliminate low amplitude samples of signal due to its high sensitivity towards the noise, and $\phi_{NL}[n]$ is the instantaneous phase non-linear component.

The standard deviation of the direct instantaneous phase non-linear component is expressed by:

$$\sigma_{dp} = \sqrt{\frac{1}{N_c} \left(\sum_{A_n[n] > A_t} \phi_{NL}^2[n] \right) - \left(\frac{1}{N_c} \sum_{A_n[n] > A_t} \phi_{NL}[n] \right)^2}, \quad (3.13)$$

where, all the parameters stay the same to the equation for σ_{ap} .

The evaluation of the spectrum symmetry, namely P , is expressed by:

$$P = \frac{P_L - P_U}{P_L + P_U}, \quad (3.14)$$

where, the P_L and P_U are denoted by:

$$P_L = \sum_{n=1}^{f_{cn}} |X_c[n]|^2, P_U = \sum_{n=1}^{f_{cn}} |X_c[n + f_{cn} + 1]|^2, \quad (3.15)$$

where, $X_c[n]$ is the expression of the Fourier transform of the signal $x_c[n]$, f_c and f_s are set to be the carrier frequency and sampling rate. Then, f_{cn} refers to,

$$f_{cn} = \frac{f_c N}{f_s} - 1. \quad (3.16)$$

The standard deviation of the absolute value of normalized and centered instantaneous amplitude is expressed by:

$$\sigma_{aa} = \sqrt{\frac{1}{N} \left(\sum_{n=1}^N A_{cn}^2[n] \right) - \left(\frac{1}{N} \sum_{n=1}^N |A_{cn}[n]| \right)^2}. \quad (3.17)$$

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

The standard deviation of the absolute value of the normalized and centered instantaneous frequency is expressed by:

$$\sigma_{af} = \sqrt{\frac{1}{N_c} \left(\sum_{A_n[n] > A_t} f_N^2[n] \right) - \left(\frac{1}{N_c} \sum_{A_n[n] > A_t} |f_N[n]| \right)^2}, \quad (3.18)$$

where, the centered instantaneous frequency f_m is the normalization of the f_N by the sampling frequency f_s according to,

$$f_N[n] = f_m[n]/f_s. \quad (3.19)$$

At this point, the referred centered instantaneous frequency f_m is denoted by the mean value of the frequency. It is given by the expression of:

$$f_m[n] = f[n] - \mu_f, \mu_f = \frac{1}{N} \sum_{n=1}^N f[n]. \quad (3.20)$$

The standard deviation of the normalized and centered instantaneous amplitude is given by the equation of:

$$\sigma_a = \sqrt{\frac{1}{N_c} \left(\sum_{A_n[n] > A_t} a_{cn}^2[n] \right) - \left(\frac{1}{N_c} \sum_{A_n[n] > A_t} a_{cn}[n] \right)^2}. \quad (3.21)$$

The kurtosis of the normalized and centered instantaneous amplitude of the received signal, is expressed by:

$$\mu_{42}^a = \frac{E \{A_{cn}^4[n]\}}{\{E \{A_{cn}^2[n]\}\}^2}. \quad (3.22)$$

The kurtosis of the normalized and centered instantaneous frequency, is expressed by:

$$\mu_{42}^f = \frac{E \{f_N^4[n]\}}{\{E \{f_N^2[n]\}\}^2}. \quad (3.23)$$

Till now, we have presented several time-domain features and transformation based features proposed in many studies. Apparently, there is exclusive usage for each individual feature.

The feature γ_{max} represents the variations in signal amplitude. On one hand, for amplitude-oriented modulations, such as QAM and amplitude-shift keying (ASK), the value of γ_{max} should be non-zero. On the other hand, for phase-oriented modulations and frequency-oriented modulations, such as PSK and frequency shift keying (FSK), the value of γ_{max} should be zero. Therefore, this feature can be used for classifications between amplitude-oriented modulations and non-amplitude-oriented modulations.

The feature σ_{aa} and σ_a reveal the variation of signal amplitude, which functions the function of γ_{max} as well. However, these features are gifted with the capability to classify the order of ASK.

The feature σ_{ap} and σ_{dp} represent the variations in signal phase. They are very useful in identifying the modulation order of PSK. Specifically, σ_{ap} is able to classify BPSK from MPSK, because there are only two possible phases in BPSK.

The feature σ_{af} represents the variations in signal frequency. Therefore, it is gifted to indicate the order of FSK and to classify PSK from FSK.

The feature P is mostly used to classify modulations with a symmetrical spectrum from modulations with asymmetrical spectrum. For instance, AM and double-sideband modulations (DSB) have symmetrical spectrums. The spectrums of vestigial sideband (VSB), lower sideband modulation (LSB) and upper sideband modulation (USB), on the contrary, are asymmetrical. Apparently, this feature is useful in analog modulations.

The feature μ_{42}^a and μ_{42}^f represent the distribution of the amplitude and the frequency, respectively. The feature μ_{42}^a is used to classify AM from ASK. Moreover, it is able to classify FSK/PSK from ASK/QAM. The feature μ_{42}^f is used to classify FM from PSK, because the frequency distribution of FM is looser, compared with FSK.

3.3.2 High-order Statistics-based Features

In 1992, applying and developing the third order moment in signal amplitude, a feature based modulation classification is proposed by Hipp, Soliman and Hsue [30]. They researched the high order moments of signal phase for MPSK classification known as common moments of modulation classification. The k th moment

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

of received signal phase in theory over the additive white Gaussian noise channel is investigated by Soliman and Hsue, which appears to be a monotonically increasing function concerning the order of *MPSK* modulation. Therefore, *MPSK* modulation for high order may give higher moment values, which makes it possible to classify *MPSK* modulation within different orders. Hence, the conclusion of effective classification for *MPSK* modulation with a higher order demanding with the higher order moment value is proposed. The expression of k th order moment of the received signal phased is given by:

$$Moment_k(r) = \frac{1}{N} \sum_{n=1}^N \theta^k(n), \quad (3.24)$$

where $\theta(n)$ stands for the phase of the n th sample of received signal.

The calculation of various k th moment of received signal $r = r[1], r[2], \dots, r[N]$ of complex valued is expressed by:

$$Moment_{xy}(r) = \frac{1}{N} \sum_{n=1}^N r^x[n] \cdot r^{*y}[n], \quad (3.25)$$

where, $x + y = k$ as well as $r^*[n]$ stands for the complex conjugate of $r[n]$.

Thereafter, a feature of the fourth order cumulant of the complex valued signal is investigated by Swami and Sadler as a solution to classify the modulation of *MPAM*, *MPSK* and *MQAM*. For instance, the second order cumulant can be designed in two different equation of:

$$C_{20} = E \{r^2[n]\}, C_{21} = E \{|r[n]|^2\}. \quad (3.26)$$

Equally, the fourth order cumulants could be defined in three definitions after applying different placements of conjugation of:

$$\begin{aligned} C_{40} &= cum(r[n], r[n], r[n], r[n]), \\ C_{41} &= cum(r[n], r[n], r[n], r^*[n]), \\ C_{42} &= cum(r[n], r[n], r^*[n], r^*[n]), \end{aligned} \quad (3.27)$$

where, $cum(\cdot)$ stands for the cumulant function. Then the joint cumulant function is expressed by:

$$cum(w, x, y, z) = E(wxyz) - E(wx)E(yz) - E(wy)E(xz) - E(wz)E(xy). \quad (3.28)$$

In the mean time, second cumulants and fourth cumulants can be calculated by the expressions of:

$$\begin{aligned}
\hat{C}_{20} &= \frac{1}{N} \sum_{n=1}^N r^2[n], \\
\hat{C}_{21} &= \frac{1}{N} \sum_{n=1}^N |r[n]|^2, \\
\hat{C}_{40} &= \frac{1}{N} \sum_{n=1}^N r^4[n] - 3\hat{C}_{20}^2, \\
\hat{C}_{41} &= \frac{1}{N} \sum_{n=1}^N r^3[n]r^*[n] - 3\hat{C}_{20}\hat{C}_{21}, \\
\hat{C}_{42} &= \frac{1}{N} \sum_{n=1}^N |r[n]|^4 - \left| \hat{C}_{20} \right|^2 - 2\hat{C}_{21}^2.
\end{aligned} \tag{3.29}$$

3.3.3 Cyclostationary Analysis-based Features

Features investigating the periodic characteristics for a cyclostationary process, namely the cyclostationary analysis of received signal is first discovered by Gardner in 1994. Based on that, Spooner, cooperating with Gardner, started to apply the cyclostationary analysis for modulation classification, while working on analyze the diversity among the spectrum appearance of different modulation types. Then, a summarization of implementation is given by Ramkuman about the cyclic feature modulation classification detection, in 2009.

For a common signal $x(t)$ of sinusoidal function, one could say it is of a cyclostationary feature or a second order periodicity if it could show its cyclic autocorrelation that is expressed in the equation of:

$$R_x^\alpha(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x\left(t - \frac{\tau}{2}\right) e^{-i2\pi\alpha t} dt, \tag{3.30}$$

where, frequency $\alpha \neq 0$. Also, the Fourier transform of the cyclic autocorrelation is then defined by:

$$S_x^\alpha(f) = \int_{-\infty}^{\infty} R_x^\alpha(\tau) e^{-i2\pi f\tau} d\tau. \tag{3.31}$$

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

Further, the cyclic spectrum could also refers to a spectral correlation function (SCF) as the definition in:

$$S_x^\alpha(f) = \lim_{T \rightarrow \infty} \lim_{\Delta \tau \rightarrow \infty} \frac{1}{T \Delta \tau} \int_{-\Delta \tau/2}^{\Delta \tau/2} X_T(t, f + \alpha/2) X_T^*(t, f - \alpha/2) dt, \quad (3.32)$$

where, contains the definition of:

$$X_T(t, \nu) = \int_{t-T/2}^{t+T/2} x(u) e^{-j2\pi \nu u} du. \quad (3.33)$$

3.3.4 Classifier

Having defined mainstream modulation features, the mainstream classifiers in modulation classification will be presented, although, some of the classification can be achieved by hierarchical decision trees. The optimization of decision thresholds is hard to implement. Therefore, machine learning has been employed for modulation classifications in the literature. On the one hand, machine learning is useful in figuring out better thresholds for decisions. Hence, machine learning based classifications can achieve higher classification rate in low SNR. On the other hand, machine learning can compress the feature space.

In this section, I briefly introduce 3 mainstream machine learning classifiers, namely artificial neural networks (ANN), support vector machines (SVM), and k-nearest neighbor (KNN).

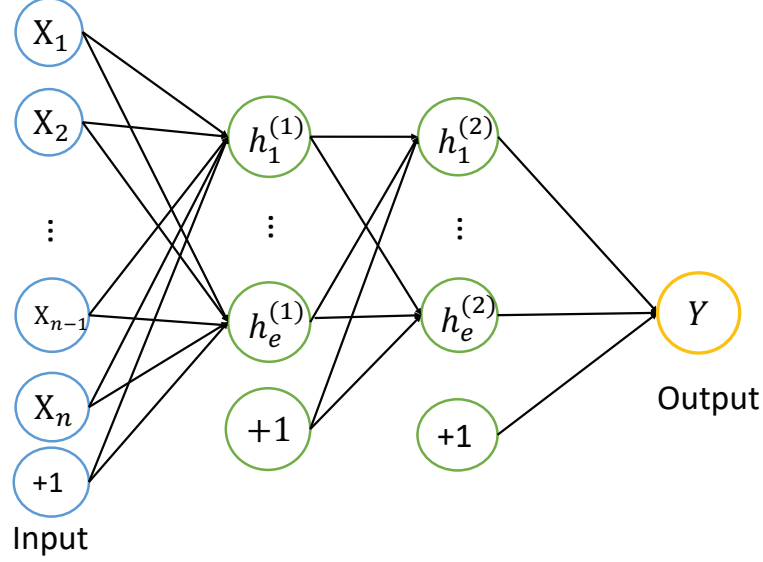


Figure 3.1: A typical multi-layer network of ANN.

- Artificial Neural Networks

Mapping features to a high-dimensional space, the neural network constructs a high-dimensional decision surface to achieve non-linear classifications. For a one-layer neural network, the output of network is identical to a linear representation of inputs, which can be given by:

$$y = W \cdot c(x) + b_1. \quad (3.34)$$

Here, b_1 is the bias parameters of the input layer. W represents the weight parameter of each node. $c(x)$ represents network inputs.

For a multi-layer neural network, the output of j -th layer can be derived by a feed-forward algorithm, given by:

$$y^j = F(W^j \cdot c_j(y^{j-1}) + b^j), \quad (3.35)$$

where, b_j is the bias parameters of the j -th layer. F is called the activation function. The nature of the neural network is a high dimensional nonlinear transformation. More nodes are equivalent to higher feature dimensions. More layers are equivalent to more non-linear transformations. A typical multi-layer network is shown in Figure 3.1.

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

As a supervised machine learning structure, the neural network needs a training phase to minimize classification error. The back propagation (BP) is widely used for this phase by calculating the change of weight and updating weight. For more information on the principles of the neural network, please refer to the Network Tutorial of Stanford University: UFLDL Tutorial.

- Support Vector Machine

In contrast to the neural network, support vector machine is achieved by reducing the dimension of the high-dimensional features. Specifically, support vector machine is useful in classification by building a hyperplane and maximizing its distance to the binary classes on each side of the hyperplane.

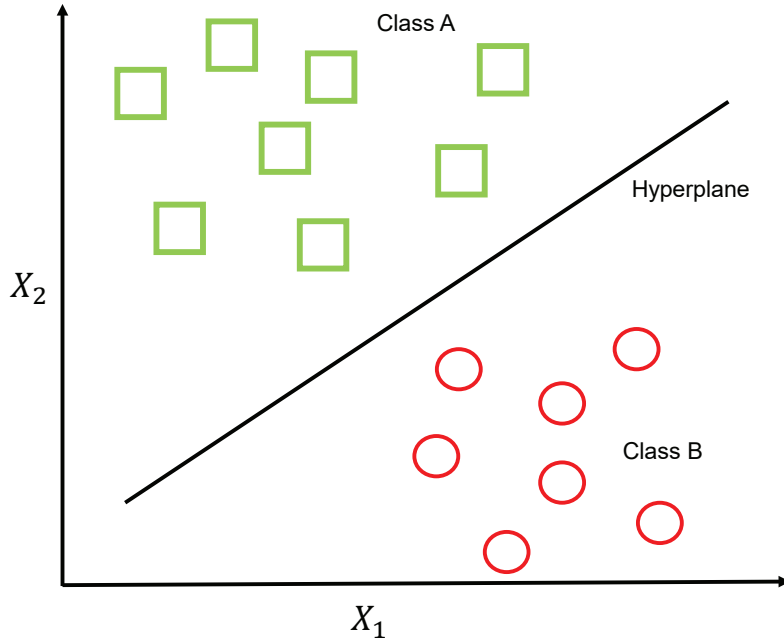


Figure 3.2: A typical two-class classification with SVM.

The linear kernels of SVM are defined by:

$$K(x, w) = x^T w, \quad (3.36)$$

where, x is the input vector, and w is the weight vector. Then the separation hyperplane is given in:

$$g(x) = x^T w + w_0, \quad (3.37)$$

where, w_0 is a bias value. w can be optimized by training phase. A typical two-class classification with support vector machine is shown in Figure 3.2. The specific training details can be found in [31].

Although ANN and its extensions are widely used for all kinds of classification tasks, deficiency in training phase reduces its classification rate. On the contrary, SVM shows superiority when the number of samples is insufficient because SVM can avoid over fitting or local minimum by dimension reduction. In general, binary SVM is mostly used in two classes of classifications. However, if more than two classes are involved in classification, SVM can be used to classify the first class from all the other classes. Then this phase should be repeated to classify the second class, and so on.

- K-Nearest Neighbor

As a non-parametric method, K-Nearest Neighbor assigns a new sample to a class by finding a K number of training base closest in distance to the new sample. Specifically, it labels the new sample with the majority of the K nearest neighbors that have same class label.

Suppose we have M reference samples for each class $M(i), i = 1, 2, \dots, I$. I is the number of classes. $F^i(m)$ represents the features of each class. F is the feature of the new sample. \hat{M} represents the samples after classification. In pseudo code, the algorithm can be expressed as follows [32]:

input:

- $M, F^i(m)$ and F

output: \hat{M}

Method:

- Step 1: Traverse the distance between F and $F^i(m)$;
- Step 2: Resort the distances matrix in descending order;
- Step 3: Select the first K distance;
- Step 4: Label all classes for all distance;
- Step 5: Mark the new sample with label i' ;

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

Step 6: Return \hat{M} as modulation result;

KNN is easy to implement because of its low complexity of calculation. However, the classification rate of KNN is severely affected by feature weight. It is necessary to standardize features before classification.

3.4 A comparison with Independent Components Analysis (ICA)

In addition to machine learning, ICA is also widely used in the field of signal classification. For example, the ICA can be used to extract individual components from a mixed signal. Therefore, ICA is often used in conjunction with arrayed signal processing techniques. In the field of modulation classification, ICA is often used for MIMO systems to obtain the signal features under the blind channel. For the extracted features, the usage of classifiers is essential [33, 34]. Therefore, ICA is more like a pre-processing process under a non-Gaussian channel. In fact, ICA and the methods described herein are not contradictory. Combining ICA with machine learning methods might effectively improve the classification performance under the ISI and other interference. For instance, a signal can be isolated from interference using ICA. A typical block diagram of MIMO modulation classification is given in Figure 3.3.

ICA can be used to separate the MIMO signal and then extract the required information from it. For the likelihood classifier, ICA can improve the accuracy of parameter estimation. For feature classifiers, ICA can provide more robust features.

3.5 Chapter Summary

In this chapter, the principle of maximum likelihood method is first presented, which is the optimal method if all information about the channel is known. The likelihood is given in equation (3.4). In fact, the maximum likelihood method is used as a reference for performance comparison, to evaluate the proposed method.

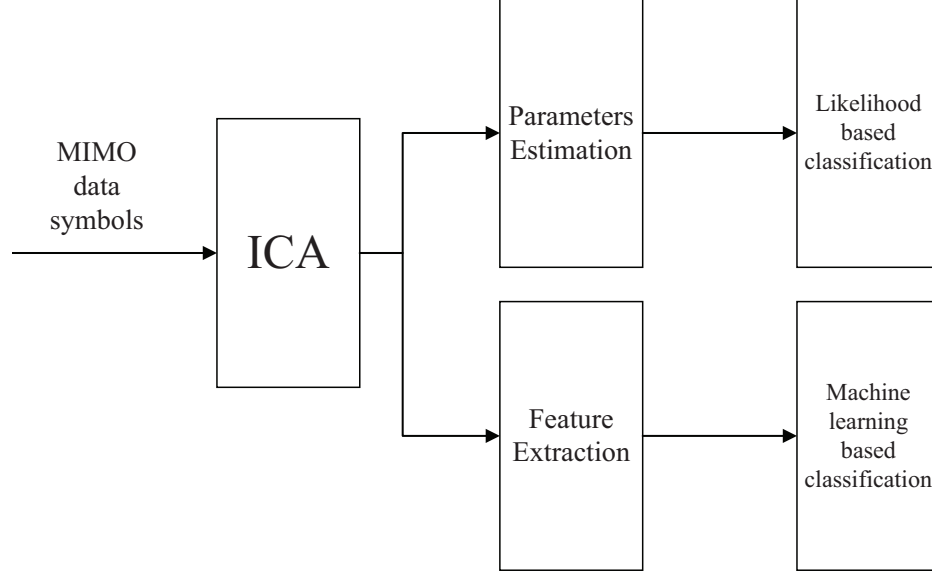


Figure 3.3: A typical block diagram of MIMO modulation classification.

Although several methods had been proposed to improve the maximum likelihood, it still maintains the dominant position on performance in ideal channel conditionals.

Hereafter, the feature-based method is introduced, which yields sub-optimal performance for much lower computational requirements. Several classical features are mainly presented, such as spectral-based features, high-order statistics-based features, and cyclostationary analysis-based features. Their suitability is analyzed for the research-related modulations. Because the high-order statistics-based features are used for one of the research, the detail of this feature will be discussed in subsequent chapters. Three mainstream classifiers are presented, namely artificial neural networks, support vector machines, and k-nearest neighbor, in the end.

3. OVERVIEW OF MODULATION CLASSIFICATION IN COGNITIVE RADIOS

Chapter 4

Supervised Machine Learning based Modulation Classification

This chapter presents a supervised machine learning based modulation classification method. Higher order cumulants are used as extracted features. Stacked denoising autoencoder (SDAE) is used as the classifier. Two scenarios have been considered in this chapter, which is rapid classification scenario and high accuracy classification scenario.

This chapter is organized as follow. Section 4.1 provides an introduction to the proposed scheme. Section 4.2 introduces the signal preprocessing and data constructions of this method. Section 4.3 presents the theory of stacked denoising autoencoder (SDAE). Then Section 4.4 presents the numerical results obtained via simulations. Section 4.5 summarizes this chapter.

4.1 Introduction

Machine learning has been used for modulation classification in many papers. [28] proposed a method based on an artificial neural network, a supervised classifier, by inputting features that contain the amplitude and phase information into an ANN network. As one of the earliest papers using machine learning for modulation classification, it has a drawback, wherein, the features it utilizes are unable to identify the scheme of the QAM that is necessary for high-speed transmission. [35] proposed a new feature to classify OFDM from QAM and BPSK, although it

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

is incompetent for more complicated modulations in single carrier systems. [26] proposed a signal recognition based on the same classifier as in this paper and it achieved an excellent performance in the AWGN channel. However, its achievement concentrates more on the recognition between the OFDM, FSK, and a few single carrier modulations in a simple scheme. This severely constrains its application in sophisticated modulation schemes that are indispensable for cognitive radio systems.

To achieve a performance improvement and QAM classification, I use a deep network instead of the ANN to achieve modulation classification with a stacked multi-layer structure. This improves the training performance of each layer. Additionally, a denoising autoencoder extends its performance by reconstructing the data from a corrupted version to extract more robust features [36]. Further, a deep network is able to discover suitable features intelligently, although these features may not be understandable [37]. Owing to the above advantages, in the past year alone, preliminary interest and discussions regarding deep learning have evolved into full-fledged conversations that have captured the attention and imagination of researchers and engineers.

4.2 Signal Preprocessing and Input Data Constructions

There are several MC methods that assume sampling synchronization, carrier recovery and waveform recovery have been achieved before the MC. However, to better understand the performance of this method, it exhibits the influence of the signal sampling synchronization over the classification ability.

This research emphasizes that although there are papers [26] [24] that utilize deep networks for signal modulations, they use pulse shaped signals rather than modulated symbols. Therefore, more input nodes are compulsory for their networks, entailing a significant calculation time. This method avoids this problem using modulated symbols.

In order to extend the application of this paper, two possible scenarios are

considered: a rapid classification scenario and a high accuracy classification scenario.

4.2.1 Rapid Classification Scenario

In certain cases, the processing time is the priority and classification accuracy can be compromised. For instance, a primary user may change its modulation from one to another frequently. Therefore, in this research, a suppose is given that a rapid classification response is more urgent than an accurate result in the rapid classification scenario. This goal can be reached using two approaches. In one approach, classifiers with low computational complexity are favored, and in the other approach, classifiers that can utilize shorter symbol sequences are preferred. Hence, every intercepted signal containing four symbols is assumed, although the classification capability may benefit from a longer input. The performance with different length of signal is not provided in this study since structure redesign is way too time consuming. In addition, complex symbols are reassigned by splitting the in-phase and quadrature symbols into two segments and line them end-to-end. These reassigned symbols are used as the input of network in this scenario.

4.2.2 High Accuracy Classification Scenario

A high accuracy classification scenario, on the contrary, ensures sufficient numbers of symbols for the input. The classification accuracy is the prior goal for this scenario. For better noise resistance, the higher-order cumulants are used as an expert feature [38], instead of raw symbols as the input of network. This is because of those cumulants, whose orders are greater than two, are zero for Gaussian random variables. In view of this, higher order cumulants are often used to compress the interference from Gaussian noise.

The n th-order cumulants of signal $x(k)$ can be represented as:

$$C_{nm} = cum[x(k), x(k), \dots, x(k), x^*(k), x^*(k), \dots, x^*(k)], \quad (4.1)$$

where n indicates the total number of $x(k)$ and $x^*(k)$, m stands for the number of $x^*(k)$, and $cum[\cdot]$ represents cumulants operation. I is used to represent

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

the set $[x(k), x(k), \dots, x(k), x^*(k), x^*(k), \dots, x^*(k)]$. Then, the n th-order cumulants operation can be represented as [39]:

$$cum(I) = \sum_{\bigcup_{p=1}^q I_p = I} (-1)^{q-1} (q-1)! \prod_{p=1}^q m(I_p), \quad (4.2)$$

where $\bigcup_{p=1}^q I_p = I$ indicates an additive operation among all the division of I . I have $q = \{1, 2, \dots, n\}$, and $p = \{1, \dots, q\}$. m is the mathematical expectation. I_p is the division of the set I . I_p satisfies $\bigcup_p I_p = I$.

The purpose of a higher-order statistic is to build a more friendly feature space for the classifier. The normalized cumulants have been calculated for the following modulation schemes: BPSK, QPSK, 8PSK, 16QAM, and 64QAM. The following cumulants are used: C_{20} , C_{21} , C_{40} , C_{41} , C_{42} , C_{60} , C_{61} , C_{62} , C_{63} , C_{80} . Their theoretical values are listed in Table 4.1. For any signal that the mean is zero, the odd-order cumulant is equal to zero [30]. Therefore, considering the potential normalization process, only the even-order cumulants are adopted. That is, only the case where n is even is considered. The data structure of the high accuracy scenario is depicted in Figure 4.1.

The work flow of the method is exhibited in Figure 4.2. The network is first trained by training data, then input the reassigned symbols or the cumulants of the intercepted signals for modulation classifications. It should be noted that all the training data used for training are pure samples that are generated without being affected by noise, timing offset or ISI channel. This research does not specifically optimize the training data for scenarios such as timing offset or ISI, although such optimizations will significantly improve the classification performance. Moreover, although multiple activation functions can be used, this study uses only the sigmoid function. This is because the network structure of this study is not very lengthy. Therefore, the problem of gradient disappearance is not the bottleneck of this study.

4.2 Signal Preprocessing and Input Data Constructions

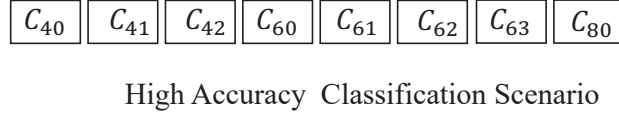
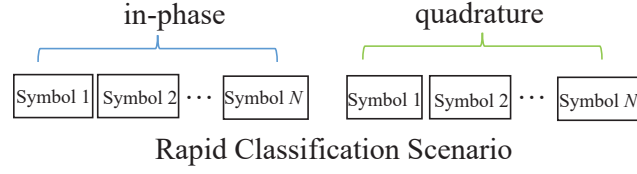


Figure 4.1: Data structures of the two scenarios.

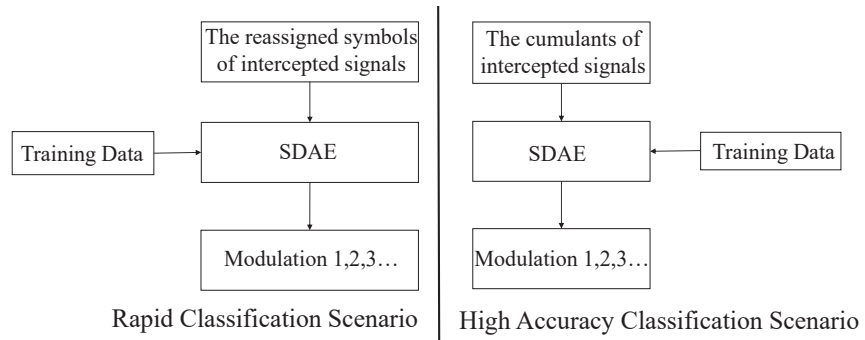


Figure 4.2: The workflow of this method.

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

Table 4.1: Higher-order cumulants of the *MPSK* and *MQAM*

	BPSK	QPSK	8PSK	16QAM	64QAM
C_{20}	1	0	0	0	0
C_{21}	1	1	1	1	1
C_{40}	-2	1	0	-0.68	-0.619
C_{41}	-2	0	0	0	0
C_{42}	-2	-1	-1	-0.68	-0.619
C_{60}	16	0	0	0	0
C_{61}	16	-4	0	2.08	1.7972
C_{62}	16	0	0	0	0
C_{63}	16	4	4	2.08	1.7972
C_{80}	-272	-34	1	-13.9808	-11.5022

4.3 Stacked Denoising Autoencoder

4.3.1 SDAE Structure

In order to appreciate the advantages of the stacked denoising autoencoder (SDAE), this section briefly reviews its structure and function. The SDAE is an extension of the sparse autoencoder, whose activation (output value) of each layer is transmitted to the next layer forward. The input layer represents the input array with a number of input units identical to the number of input features. The hidden layers, whose values are not observed in the training set, contain multiple layers with parameters that are obtained by greedy layer-wise training [40]. The word stacked indicates a different way of training that pre-trains a deep network for

each layer in turn. Compared with a conventional neural network, the advantages of SDAE are as follows:

- Use drop out at the input level to improve recognition accuracy.
- Layer-wise pre-training is used to avoid falling into the local minimum trap.

Therefore, SDAE can achieve higher classification accuracy.

The output layer consists of a softmax classifier that is capable of classifying the modulation as desired. The softmax classifier is a supervised classifier that regards its output vector as a probability distribution over candidate modulations. It can maximize the conditional probability of a sample corresponding to its correct class [41]. The number of output units is identical to the number of modulation schemes that is desired to classify. The autoencoder is trained by a reconstructed input from its corrupted version. The structure of the SDAE in this research is displayed in Figure 4.3. The network structure is decided using the training success rate. Specifically, the network is trained by a data base and then, test the success rate of the training by different data. The number of nodes is changed in the 1st hidden layer to determine the best number of nodes that have the highest success rate of training. Then, the structure of the 1st hidden layer is set. This process is repeated to determine the best number of nodes for the 2nd hidden layer.

The specific mathematical explanations are given by subsequent sections.

4.3.2 SDAE Forward Propagation

The autoencoder tries to learn a function that aims at transforming its input, x , into an output, \hat{x} , similar to x [42]. It values its distortion by the cost function and determines the optimized activation of each node. Let x be the original input array of the network. According the definition of denoising autoencoder [43], the real input array of the network is $c(x)$. Here, c is a non-deterministic procedure of corruption. Then the values of the hidden layer can be obtained by:

$$y = F(W \cdot c(x) + b_1), \quad (4.3)$$

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

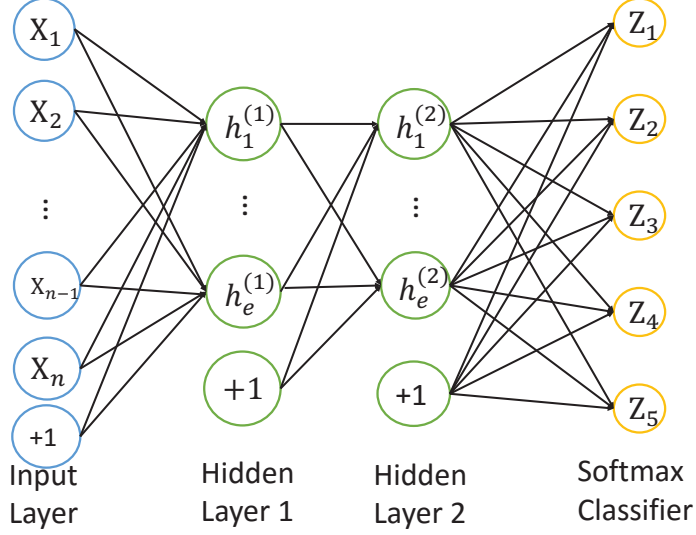


Figure 4.3: Structure of the SDAE with two hidden layers.

and the output of the hidden layer is,

$$z = F(W^T \cdot y + b_h). \quad (4.4)$$

Here, F is called the activation function. In most cases, the activation function is chosen to be a sigmoid function or a hyperbolic tangent function. b_1 and b_h are the bias parameters of the input layer and the hidden layer. W represents the weight parameter of each node.

A stacked architecture uses the output of a hidden layer as the input of the next hidden layer. Let x^j , y^j , and z^j represent the input, value, and output of a hidden layer, j , respectively. Then, the forward propagation of the entire SDAE can be expressed by:

$$y^j = F(W^j \cdot c_j(y^{j-1}) + b_h^j). \quad (4.5)$$

Let L stand for the last hidden layer of the network. Then, the final output of SDAE with a softmax classifier can be expressed by:

$$y^L = \text{softmax}(W^L \cdot y^{L-1} + b_h^L). \quad (4.6)$$

Here, b_h^L is the bias parameter of the last hidden layer.

The forward propagation decides the relations between the input and output of the network. However, the difference between the input and output needs to

be minimized to obtain the correct weight parameters of each nod. This goal can be achieved by the SDAE fine-tuning.

4.3.3 SDAE Fine-tuning

A set of training examples are assumed, $\{x^1, \dots, x^m\}$. These examples belong to K classes. $k = \{1, 2, \dots, K\}$ is given here. The network outputs and the cost function of a single training example are defined by $h_{W,b}(x)$ and $J(W, b; x)$, respectively. Then, the cost function of a single example can be given by:

$$J(W, b; x) = \frac{1}{2} \|h_{W,b}(x) - x\|^2, \quad (4.7)$$

where W and b represent the weight and the bias parameters of the hidden layers, respectively. For a training set of m examples, the average error term of the overall cost function can be defined by:

$$J(W, b) = \frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}), \quad (4.8)$$

where m is the number of examples, $x^{(i)}$ represents the input vectors. To compress the magnitude of the weights and prevent overfitting, a decay has to be applied to the weight terms that transforms (4.7) into:

$$J(W, b) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{W,b}(x^{(i)}) - x^{(i)}\|^2 \right) + \frac{\lambda}{2} \sum_{l=0}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2, \quad (4.9)$$

where λ is the weight decay parameter for controlling the relative weight of the two components. n_l is the number of layers. s_l denotes the number of nodes in layer l . $W_{ij}^{(l)}$ denotes the parameter between unit j of layer l and unit i of layer $l+1$. As the sparse autoencoder embraces a spare constraint, most of the hidden units may retain zero, if a sigmoid activation function is applied here. Then, the penalty term can be expressed by:

$$KL(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}, \quad (4.10)$$

which is based on the Kullback-Leibler divergence [36], a function to measure the difference between two distributions. $\hat{\rho}_j$ is the average activation level of the j th

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

hidden unit for all the data. ρ is the sparsity parameter and is typically a small value around zero.

Considering this, it is apparent that the overall cost function of sparse autoencoder with the weighted sparse penalty term is given by:

$$J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} KL(\rho \parallel \hat{\rho}_j), \quad (4.11)$$

where s_2 represents the number of hidden units and β controls the weight of the sparsity penalty term. Considering the sparse autoencoder is followed by a softmax classifier, the final output of entire network is a vector that contains K elements. Therefore, the final cost function of the entire network is:

$$J(W, b) = -\frac{1}{m} \sum_{i=1}^m \left(\log \frac{e^{y_{t_i}^{(L)}}}{\sum_{k=1}^{S_L} e^{y_k^{(L)}}} \right) + \frac{\lambda}{2} \sum_{l=0}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2. \quad (4.12)$$

Here, t_i stands for the index corresponding to the correct label for sample i . To optimize the activation of each layer, $J(W, b)$ should be minimized as a function of W and b . Changing W and b , the minimum value of $J(W, b)$ can be found. A typical way to figure out the minimum value of $J(W, b)$ is to use the gradient descent [44].

4.4 Simulation and Results

To train the classifier, a training database with 10,000 samples for each modulation had to be built in advance. The activation function is a sigmoid function. I intercepted four symbols for the rapid classification scenario and 500 symbols for the high accuracy scenario. The parameters used for SDAE (training and classification) are shown in Table 4.2.

4.4.1 Rapid Classification Scenario

The network structure for the rapid classification scenario is an 8-64-32-5 structure. This structure was selected by a comparison of the training error. To determine the number of nodes to be used in each layer, the change in training

Table 4.2: Parameters used for SDAE (training and classification)

Parameter	Value
Input layer corruption	0.4
1st hidden layer corruption	0.4
1st hidden layer sparsity	0.05
2nd hidden layer sparsity	0.05
Number of training samples	10,000
Learning rate	1
Activation function	sigmoid

error corresponding to different numbers of nodes in each layer are depicted, as shown in Figure 4.4. We design the structure of SDAE layer by layer because of the layer-wised pre-train. It is a reasonable inference that a layer-wised structure design could improve the pre-training effect. The top plot is for the 1st hidden layer and the bottom plot is for the 2nd hidden layer. Apparently, 64 nodes of the 1st layer and 32 nodes of the 2nd layer achieve the highest probability of a successful training. Therefore, an 8-64-32-5 network structure was selected.

The execution time is given in Table 4.3. Note that once the training phase is finished, SDAE does not need to repeat this phase for each test signal realization. Therefore, the time cost of training phase should not be included in the time cost of the classification process.

The same hardware environment was used for both methods. Because of the high complexity of its algorithm, ML clearly requires more time to process than SDAE. The method has an advantage over ML with respect to running speed (14 times faster). Moreover, the time consumed by the SNR estimation for ML

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

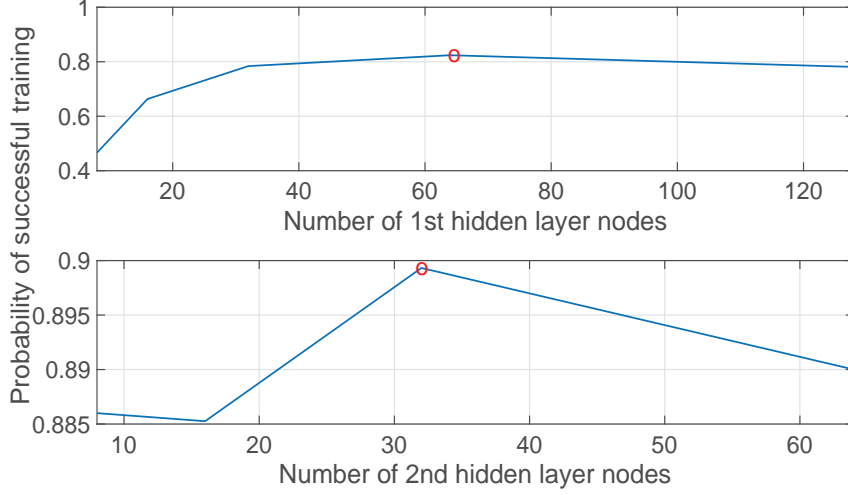


Figure 4.4: Change in the training error corresponding to the different numbers of nodes in each layer.

Table 4.3: Execution time comparison of proposed method with ML (10,000 iterations)

	SDAE (classification part)	ML	SDAE (training part)
Execution Time (s)	1.001279	14.465184	32.886

cannot be ignored, because all the channel parameters have to be known by ML. Therefore, one can conclude that the proposed method is more competent than ML because classification speed is the priority of this scenario.

In Figure 4.5, the classification accuracy of proposed method is compared with that of ML [19] for the rapid classification scenario. For each value of the SNR, 10,000 realizations of the test data were produced. All five modulation types were considered simultaneously.

Because the ML classifier obtains optimal performance under ideal conditions if expert features are not utilized, it is more accurate than my method at low SNRs. The ML shows a P_{cc} of $> 97\%$ at 20 dB. The method exhibits a P_{cc} of

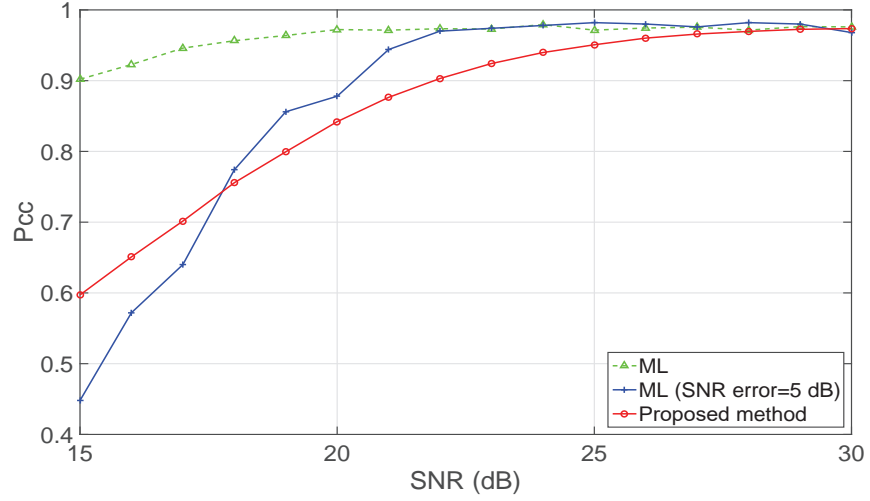


Figure 4.5: Classification accuracy comparison of the proposed method with ML.

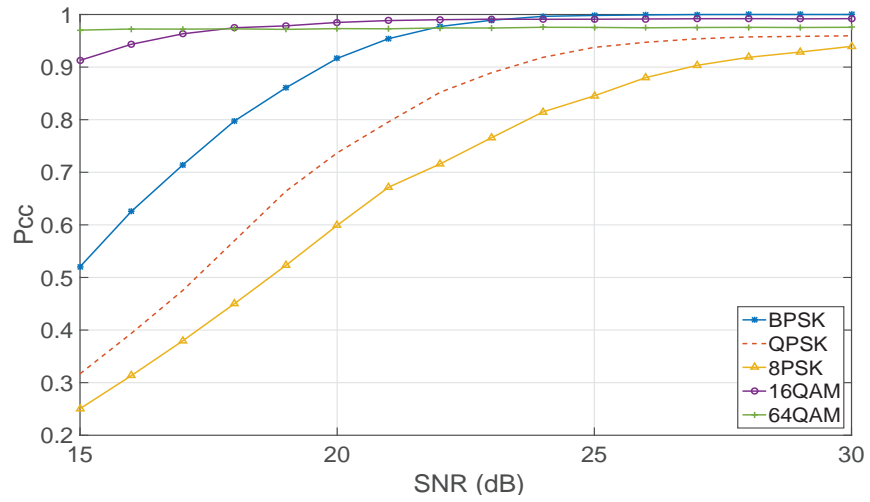


Figure 4.6: Classification accuracy for different individual modulations.

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

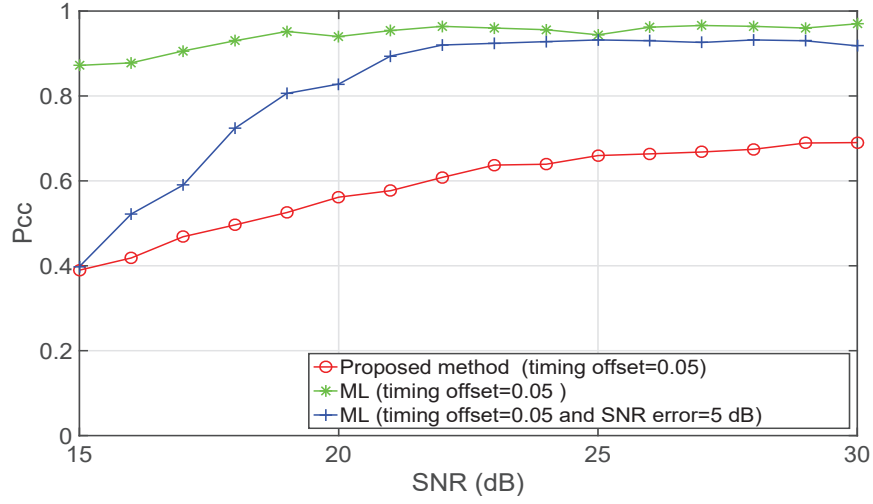


Figure 4.7: Influence of the signal sampling synchronization on classification ability.

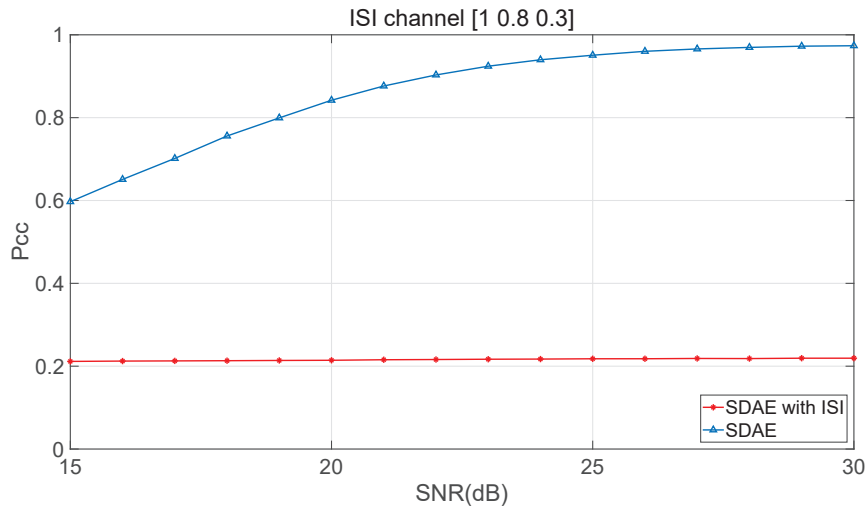


Figure 4.8: An ISI channel with a parameter of [1 0.8 0.3] is considered.

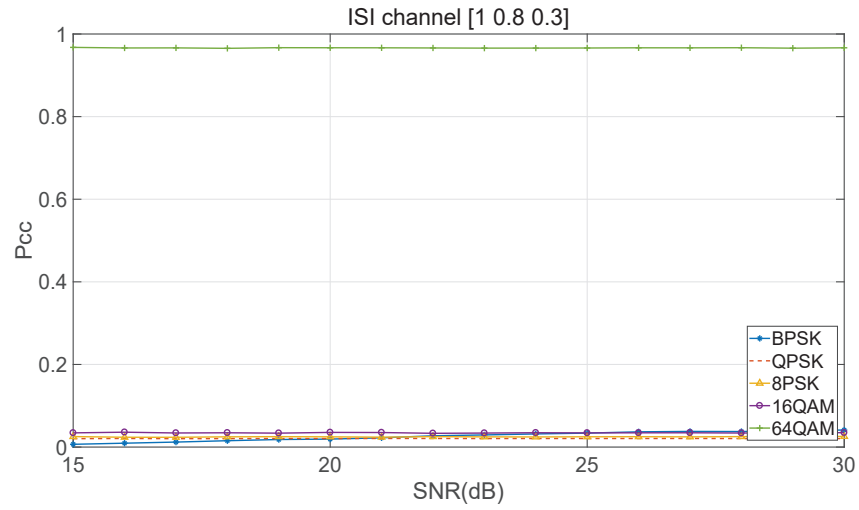


Figure 4.9: Classification accuracy for different individual modulations under ISI.

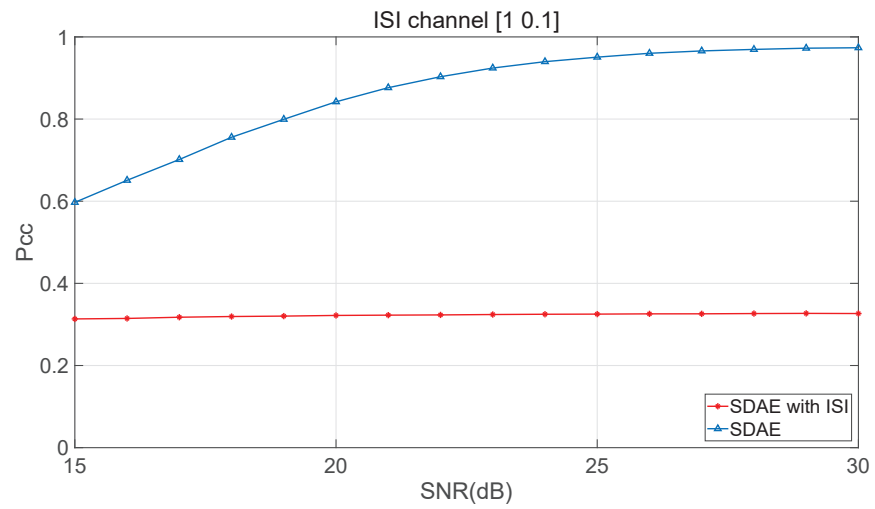


Figure 4.10: An ISI channel with a parameter of [1 0.1] is considered.

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

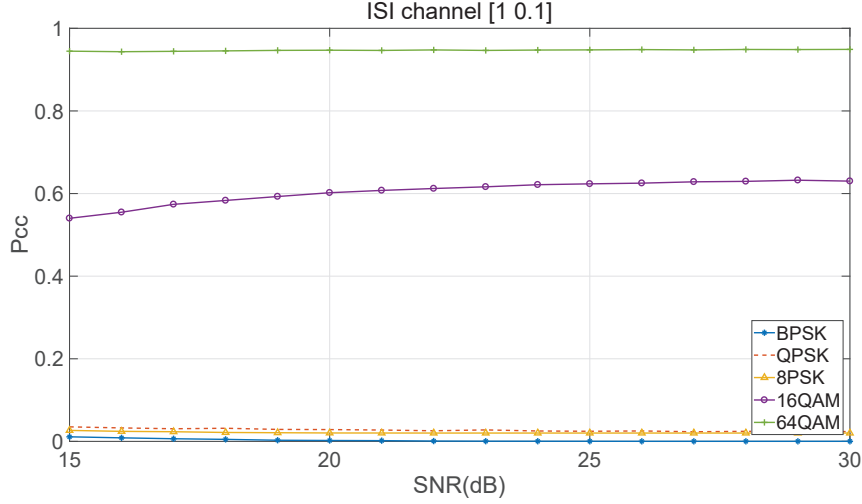


Figure 4.11: Classification accuracy for different individual modulations under ISI.

> 97% at 29 dB. However, the system requirements of the ML method are much stricter than those of the method. One of those requirements is that the noise variance of the intercepted signal has to be known. In the modulation classification field, this requirement is an impractical assumption in most cases. Therefore, the SNR estimation has to be applied. To date, there are few SNR estimation methods that work for modulation classification. Most SNR estimation methods are data-aided [45], which means that a perfect knowledge or an estimate of the transmitted sequence is necessary for the estimators. In addition, the limitation of symbol length is another factor that cannot be ignored. Therefore, the ML cannot practically work ideally without any performance degradation. To fairly compare the ML method and the proposed method, the influence of SNR error was considered. The SNR error in Figure 4.5 is 5 dB. The ML shows a P_{cc} of > 97% at 24 dB with SNR error. Apparently, the accuracy gap between the ML and the proposed method is narrowed. Moreover, the P_{cc} of the ML with SNR error degrades faster than the proposed method when SNR decreases. Certainly, the SNR error should be much smaller than 5 dB in practical use, hence the performance gap between ML and proposed method could be enlarged. This setting is only for a better insight into the potential of the ML method.

The performance for different individual modulations is given in Figure 4.6.

It is clear that a lower SNR degrades classification accuracy for all modulations. All of the modulations, 16QAM and 64QAM have higher classification accuracy than MPSK. BPSK has higher classification accuracy than QPSK and 8PSK at lower SNRs.

In Figure 4.7, the influence of the signal sampling synchronization is investigated over the classification ability. Although timing offset is always time-varying in practical use, a fixed timing error of one sample was considered. As the raised cosine transmit filter was configured to have 20 output samples per symbol, the timing offset was 0.05. Clearly, both methods are affected by the timing offset. SDAE is more sensitive to timing offset because the network enlarges the distortion caused by it. Incidentally, both the methods fail when a higher timing offset exists. Therefore, timing offset should be compensated for to obtain better classification accuracy.

ISI channels have been considered as well. From Figure 4.8 to Figure 4.11, the influences of ISI are investigated over the classification ability. In Figure 4.8, an ISI channel with a parameter of $[1 \ 0.8 \ 0.3]$ is considered. Apparently, proposed method cannot function under this channel. This is expected because the second and the third delay cannot be ignored.

The performance for different individual modulations under ISI is given in Figure 4.9. It is clear that modulation other than 64QAM can not be correctly classified. To be precise, all the modulation is classified as 64QAM. This is because the ISI has caused significant interference with the amplitude of all the modulated signals.

In Figure 4.10, an ISI channel with a parameter of $[1 \ 0.1]$ is considered. This time, the proposed method still cannot accurately classify all the modulation methods under this channel, although the second delay is relatively much smaller than the previous scenario.

The performance for different individual modulations under this ISI is given in Figure 4.11. It is clear that only 64QAM can be correctly classified. Moreover, 16QAM can be classified occasionally because 16QAM also contains a variety of amplitude information.

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

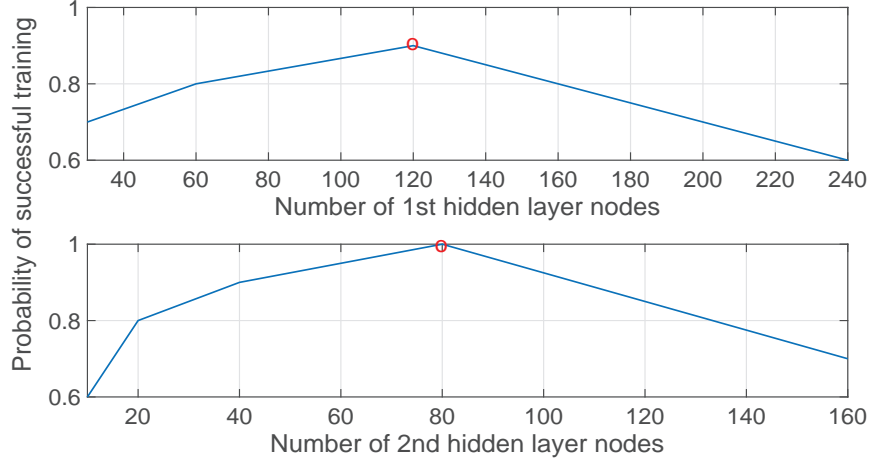


Figure 4.12: Change in the training error corresponding to the different numbers of nodes in each layer.

4.4.2 High Accuracy Classification Scenario

The network structure for the high accuracy classification scenario is an 8-120-80-5 structure. This structure was selected by a comparison of the training error. To determine the number of nodes to be used in each layer, the change in the training error corresponding to the different numbers of nodes in each layer is depicted in Figure 4.12. The top plot is for the 1st hidden layer and the bottom plot is for the 2nd hidden layer.

Apparently, 120 nodes of the 1st layer and 80 nodes of the 2nd layer display the highest probability of a successful training. Therefore, an 8-120-80-5 network

Table 4.4: Execution time comparison of proposed method with ML and KNN (10,000 iterations)

	SDAE (proposed method)	ML	KNN
Execution Time (s)	110.939827	950.636033	548.826936

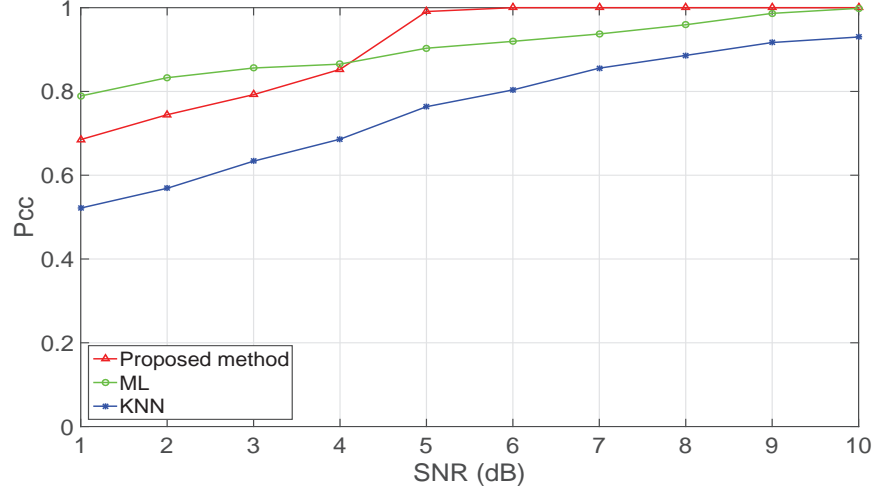


Figure 4.13: Classification accuracy comparison of our method with ML and KNN.

structure was selected .

The execution time is given in Table 4.4. the same hardware environment was used for all methods. ML clearly requires the most time to process than proposed method, although processing time is not the priority in this scenario. The proposed method has an advantage over ML with respect to running speed (9 times faster). Moreover, the proposed method has an advantage over KNN with respect to running speed (5 times faster). Therefore, one can conclude that the proposed method is more efficient than ML and KNN.

The performance of proposed method is compared with the k-nearest neighbor [23] and ML [19] methods for the high accuracy scenario in Figure 4.13. For each value of the SNR, 10,000 realizations of the test data had been produced. The ML shows a P_{cc} of $> 97\%$ at 9 dB. The proposed method shows a P_{cc} of $> 97\%$ at 5 dB. The superior performance of this method is reasonable because it can be interpreted as an integration of the higher-order cumulants, sparse autoencoders, and a softmax classifier. The higher-order cumulants suppress Gaussian noise [38]. The autoencoder extracts feature from the higher-order cumulants [42]. Then, the softmax classifier achieves maximum likelihood classifications [46]. The proposed method performs better than KNN, although KNN also utilizes higher-order cumulants. The accuracy degradation of my method at 4 dB is caused

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

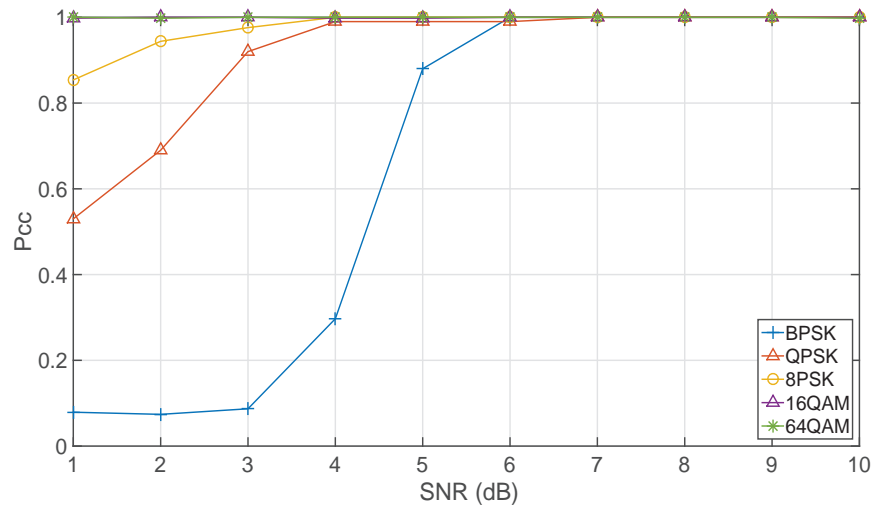


Figure 4.14: Classification accuracy for different individual modulations.

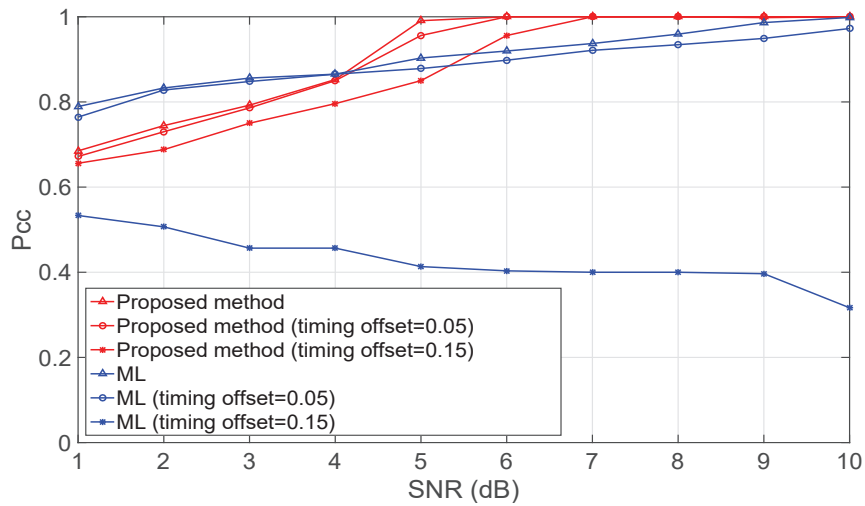


Figure 4.15: Influence of the signal sampling synchronization on classification ability.

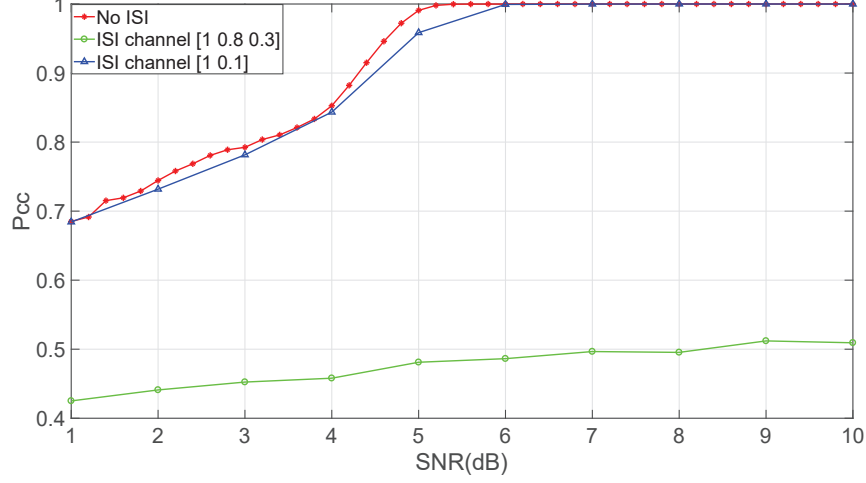


Figure 4.16: Influence of the ISI on classification ability.

by the BPSK. This can be observed if the performance for different individual modulations is investigated in Figure 4.14. It is clear that a lower SNR degrades classification accuracy for all modulations. Of all the modulations, 16QAM and 64QAM have higher classification accuracy than *MPSK*. BPSK has the lowest classification accuracy at lower SNRs.

The influence of the signal sampling synchronization over the classification ability is shown in Figure 4.15. A timing error of one sample was supposed. The raised cosine transmit filter sampled 20 times per symbol. Therefore, the timing offset was 0.05. The proposed method shows a P_{cc} of $> 97\%$ at 6 dB. The ML exhibits a P_{cc} of $> 97\%$ at 10 dB. Next, a timing error of three samples is supposed. The timing offset is 0.15. The proposed method has a P_{cc} of $> 97\%$ at 7 dB. The ML cannot function under this timing offset. Clearly, the proposed method is slightly affected by the timing offset. This robustness to timing offset is attributable to the higher-order cumulants. On the contrary, ML cannot function as usual when the timing offset increases. This is expected because ML requires all channel parameters to be known. A severe timing offset leads to a mismatch between the model used in the classifier design and the actual statistics in the channel. Therefore, one can conclude that the proposed method is superior to the ML when sampling is not synchronized.

ISI channels have been considered as well. In Figure 4.16, the influences of

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

ISI are investigated over the classification ability. Specifically, an ISI channel with a parameter of $[1 \ 0.8 \ 0.3]$ and an ISI channel with a parameter of $[1 \ 0.1]$ are considered. Apparently, the performance of proposed method has been seriously affected by the first ISI channel. This is expected because the second and the third delay cannot be ignored. However, the proposed method can accurately classify all the modulation methods under the second ISI channel, because the second delay is relatively much smaller than the previous scenario.

4.5 Chapter Summary

Table 4.5: Summary of the three methods

	Rapid Classification Scenario	High Accuracy Classification Scenario
SDAE	High execution speed Sensitive to timing offset	High classification accuracy Robust to timing offset
ML	Low execution speed Sensitive to timing offset	Medium classification accuracy Sensitive to timing offset
KNN	Unsuitable for this scenario	Lowest classification accuracy

4.5.1 Rapid Classification Scenario

First, the proposed method has an advantage over ML with respect to running speed (14 times faster). Moreover, the time consumed by the SNR estimation for ML cannot be ignored. Although the classification accuracy degrades fast when a timing offset exists, it is less complex to accomplish the timing offset estimation and recovery [47] than the blind SNR estimation [48]. Therefore, one

can conclude that the proposed method is more competent than ML because classification speed is the priority of this scenario.

Second, the KNN cannot be used in the rapid classification scenario because a long sequence of symbols is necessary for it to calculate the cumulants.

4.5.2 High Accuracy Classification Scenario

First, the proposed method is more accurate than ML in the high accuracy classification scenario because this method integrates the advantages of the higher-order cumulants and SDAE. The ML shows a P_{cc} of $> 97\%$ at 9 dB. The proposed method shows a P_{cc} of $> 97\%$ at 5 dB. This advantage is further increased if a timing offset exists. It shows a P_{cc} of $> 97\%$ at 6 dB while timing offset is 0.05 symbols. The ML exhibits a P_{cc} of $> 97\%$ at 10 dB while timing offset is 0.05 symbols.

In addition, the proposed method obtains a higher classification accuracy than the KNN in the high accuracy scenario. The KNN shows a P_{cc} of $> 78\%$ at 5 dB. The proposed method shows a P_{cc} of $> 97\%$ at 5 dB. It shows superiority to KNN in both the scenarios.

4.5.3 Discussion

A summary of the advantages and disadvantages of the proposed method, ML, and KNN is given in Table 4.5.

The performance of proposed method can be improved if more time could be spent on the optimization process. The goal is to propose a digital modulation classification scheme for CR systems. Therefore, an analysis of the relations between the number of cumulants and classification accuracy was not performed. A higher or lower number of cumulants could affect the classification accuracy.

In addition, complex symbols are used rather than pulse shaped complex signals as the network input, simplifying the network topology and reducing the calculation overhead. This can be easily verified if our network structure is compared with other convolution neural network-based methods [24] and conventional neural network-based methods. [24] trains its network for 23 minutes

4. SUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

over a 900,000-sample training set. The time it consumes for the training phase is approximately 46 times that of the proposed method.

Drawbacks for this method are inevitable. First, the training phase may consume considerable time, although it can be finished in advance. Using a platform with higher calculation performance may solve this problem. In computer vision research, graphics processing units have been widely used to speed up the process of the training phase. This could be a feasible solution to the high calculation overhead. Next, as long as the length of the input signal is changed, the network structure has to be redesigned for a rapid classification scenario. This could be a time-consuming portion of this method. The use of expert features could be a feasible solution because the number of features can be decided in advance.

Chapter 5

Unsupervised Machine Learning based Modulation Classification

This chapter presents an unsupervised machine learning based modulation classification method. Time-frequency analysis is used for feature extractions. Density-based spatial clustering of applications with noise is used as the classifier. The proposed method shows a stronger ability of classification than conventional methods.

This chapter is organized as follow. Section 5.1 provides an introduction to the proposed scheme. Section 5.2 introduces the feature extraction of this method. Section 5.3 presents the theory of DBSCAN and the way of parameter decisions. The classification scheme is presented in section 5.4. Then section 5.5 presents the numerical results obtained via simulations. Section 5.6 summarizes this chapter.

5.1 Introductions

Many studies utilized time-frequency distributions to perform modulation classification. In [49], the authors proposed a method based on the Margenau-Hill Distribution (MHD), which is a time-frequency distribution. It uses MHD to extract amplitude and phase features for the subsequent classifier. However, it works only for PSK signals and 8QAM signals. Moreover, some of the parameters that it uses are decided based on experience, and not scientific derivation. Further, its performance in low signal-to-noise ratio conditions cannot satisfy the

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

demand in real applications. In [50], the authors proposed a method based on the pseudo Wigner-Ville distribution (PWVD), which is the same distribution as that used in this paper. It utilized successive spectral slots to identify the number of carrier frequencies. Therefore, it is able to distinguish FSK signals from PSK signals. However, because it cannot classify signals with different orders, it has only limited use in practice. In [35], the authors modified the method in [50] by using the maximum power spectral density (PSD). It can classify FSK signals in different orders, but not PSK signals with different orders.

The contributions and novelties of this method are as follows:

To better understand what can be done in the time frequency distribution (TFD), a general model is first considered for the signal and the definitions of the relevant modulations.

5.2 Feature Extraction

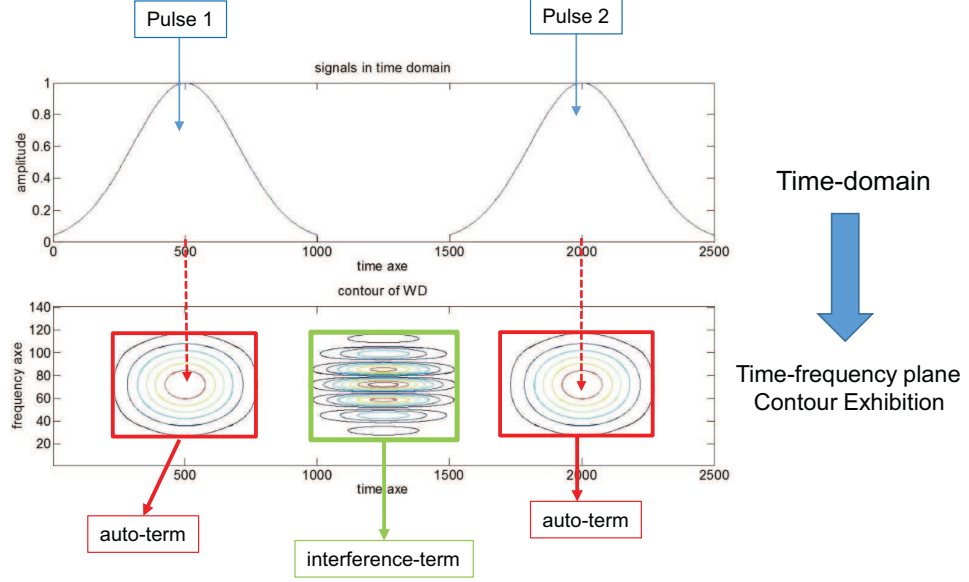
5.2.1 Feature-Extraction Tool

The Wigner-Ville distribution (WVD) is a type of TFD that analyses signals in the time-frequency domain. The TFD is a view of a signal (taken to be a function of time) represented over both time and frequency [51], and it is often complex-valued over time and frequency, e.g., the short-time Fourier transform (STFT). One form of the TFD can be formulated by multiplying a signal by its time delay. This formulation was proposed by Eugene Wigner in 1932, and it was modified by Ville in 1948 [52]. To suppress the interference terms, a smoothing function is used in order to execute a convolution because interference terms are oscillatory [53]. The pseudo-WVD (PWVD), which is a form of smoothed WVD, is defined as

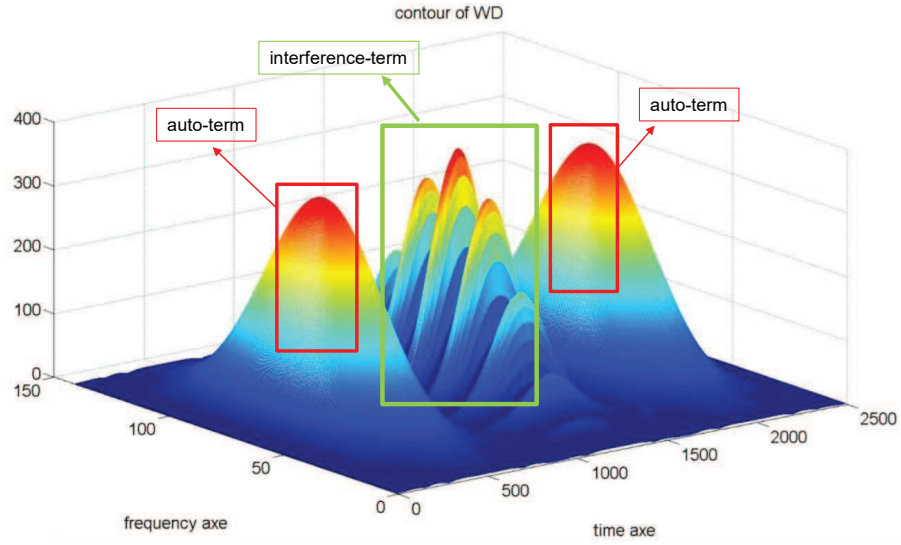
$$W_x(t, f) = \int_{\tau} x(t - \frac{\tau}{2})x^*(t + \frac{\tau}{2})h(\tau)e^{-2j\pi f\tau}d\tau, \quad (5.1)$$

where τ is the time delay, f is the frequency and $h(\tau)$ is a smoothing function.

Figure 5.1 shows a signal that contains two pulses in the time domain, and its contour on time-frequency plane. In Figure 5.1(a), the upper figure represents a signal that contains two pulses in the time domain. Moreover, there is no



(a) A signal that contains two pulses in time domain



(b) 3D vision of signal contour on time-frequency plane

Figure 5.1: A signal that contains two symbols in time domain, its contour and the 3D vision on time-frequency domain (PWVD).

overlap between pulses. The lower figure in Figure 5.1(a) is the contour exhibition of the time-frequency plane. Specifically, the auto-terms with red rectangles

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

are produced by pulses in the time domain. Between auto-terms, a component named interference-term appears. This component is generated by the coexistence of pulses. Unlike the short-time Fourier transform, PWVD is a bilinear time-frequency distribution. Therefore, any two independent components produce an interference-term. In most cases, scholars try as far as possible to eliminate this interference-term. In this study, however, it is the key to extract phase features. Different phase shifts lead to different initial oscillation phases, which in turn lead to interference-term shifts along the frequency axis. The 5.1(b) shows the 3D time-frequency distribution of the same signal. An obvious oscillation can be found from the interference term.

According to [51], the envelop of interference-term is identical with those of the auto-terms. The coordinate of interference-term depends on the coordinate of auto-terms. The interference-terms oscillate along the frequency axis. In particular, the further that two signals are, the more rapid the interference-terms oscillate. To better understand these characteristics, we turn to mathematical derivations.

Assuming that $s_1(t), s_2(t)$ are two signals, then the cross-PWVD is given by

$$W_{s_1, s_2}(t, f) = \int_{\tau} s_1(t - \frac{\tau}{2}) s_2^*(t + \frac{\tau}{2}) h(\tau) e^{-j2\pi f \tau} d\tau. \quad (5.2)$$

The auto-PWVD of a signal s_1 is defined as

$$W_{s_1}^{Auto}(t, f) = W_{s_1, s_1}(t, f). \quad (5.3)$$

From the PWVD's hermiticity property $W_{s_1, s_2}(t, f) = W_{s_2, s_1}^*(t, f)$, one could know that the auto-PWVD of signal s_1 has real values.

Assume a pulse-shaped symbol:

$$x_i(t) = A_i g(t). \quad (5.4)$$

The complex symbol and Nyquist pulse are A_i and $g(t)$, respectively. The auto-PWVD of a pulse shaped symbol is defined as

$$W_{x_i}^{Auto}(t, f) = W_{x_i, x_i}(t, f) = \int x_i(t + \frac{\tau}{2}) x_i^*(t - \frac{\tau}{2}) h(\tau) e^{-j2\pi f \tau} d\tau. \quad (5.5)$$

According to (5.4), (5.5) can be rewritten as

$$W_{x_i}^{Auto}(t, f) = |A_i|^2 W_0(t, f), \quad (5.6)$$

where $W_0(t, f)$ is defined as

$$W_0(t, f) = \int_{\tau} g(t - \frac{\tau}{2}) g^*(t + \frac{\tau}{2}) h(t) e^{-2j\pi f\tau} d\tau. \quad (5.7)$$

As the modulated signal contains multiple symbols, it can be thought of as a multi-component signal. An N-component signal $s(t)$ is considered as,

$$s(t) = \sum_{i=1}^N x_i(t). \quad (5.8)$$

The auto-PWVD of the N-component signal $s(t)$ can be expressed as

$$W_s^{Auto}(t, f) = \sum_{k=1}^N \sum_{l=1}^N W_{x_k x_l}(t, f) = \sum_{k=1}^N W_k^S(t, f) + \sum_{k=1}^N \sum_{\substack{l=1 \\ l > k}}^N W_{kl}^I(t, f), \quad (5.9)$$

where

$$W_k^S(t, f) = W_{x_k}^{Auto}(t, f) \quad (5.10)$$

and

$$W_{kl}^I(t, f) = W_{x_k^*, x_l}(t, f) + W_{x_k, x_l^*}(t, f) = 2Re \{ W_{x_k, x_l^*}(t, f) \} \quad (5.11)$$

are PWVD signal term corresponding to the k -th component and PWVD interference term corresponding to the k -th and l -th components, respectively.

5.2.2 Amplitude-Feature Extraction

To discriminate between MQAM and MPSK, one have to go back to (5.6). Every shaped symbol is made as an input of PWVD. Let $f = f_c$ and $t = T_s/2$. Then, (5.6) becomes

$$W_i^{Auto}(\frac{T_s}{2}, f_c) = |A_i|^2 W_0(\frac{T_s}{2}, f_c), \quad (5.12)$$

where $W_0(T_s/2, f_c)$ can be treated as a constant. According to (5.12), it is clear that $|A_i|^2$ is the amplitude feature that is extracted from PWVD, which represents

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

the symbol energy. For *MPSK*, the components of the in-phase and quadrature are $A\cos(\theta_i)$ and $A\sin(\theta_i)$, respectively. θ_i is the polar angle of A_i . Then, the symbol energy is $A^2\cos^2(\theta_i) + A^2\sin^2(\theta_i) = A^2$. Alternatively, for *MQAM*, if the components of the in-phase and quadrature are I_i and Q_i , respectively, then the amplitude factor of PWVD is $I_i^2 + Q_i^2$. As *MPSK* has no direct amplitude information, *MPSK* generates the same symbol energy. However, because of the difference in symbol energy, 64QAM contains more kinds of energy levels of the symbol than 16QAM, which can be revealed by clustering the amplitude information extracted from PWVD.

5.2.3 Phase-Feature Extraction

Let $x_1(t)$ and $x_2(t)$ represent two pulse-shaped symbols that are time shifted from $x_0(t)$. A_1 and A_2 represent complex symbols.

$$x_1(t) = A_1x_0(t - t_1), \quad (5.13)$$

$$x_2(t) = A_2x_0(t - t_2). \quad (5.14)$$

Assume that $t_1 = 0$. Then, t_2 can be expressed by T_s , which is the symbol period. The auto-PWVD is applied to the signal $s(t) = x_1(t) + x_2(t)$. The signal term of $x_1(t)$ can be expressed as

$$W_1^{Auto}(t, f) = |A_1|^2 W_0(t, f). \quad (5.15)$$

According to the property of the time-frequency shift invariance, the signal term of $x_2(t)$ can be expressed as

$$W_2^{Auto}(t, f) = |A_2|^2 W_0(t - T_s, f). \quad (5.16)$$

Meanwhile, the interference term is given by [51]

$$W_{12}^I(t, f) = 2|A_1A_2|W_0(t - \frac{T_s}{2}, f)\cos[\Theta], \quad (5.17)$$

$$\Theta = 2\pi(T_sf) + \phi_{12}, \quad (5.18)$$

$$\phi_{12} = \arg A_1 - \arg A_2. \quad (5.19)$$

(5.17) indicates the oscillation property of interference terms. Because A_1 and A_2 are two complex symbols, the following equation can be employed for phase extraction:

$$\phi_{12} = \arg A_1 - \arg A_2 = \theta_1 - \theta_2 = \Delta\theta. \quad (5.20)$$

where θ_1 and θ_2 are polar angles of A_1 and A_2 , respectively. Equation (5.20) indicates that the phase shift of adjacent symbols can be replaced by the oscillation phase of interference terms.

To extract the interference term's oscillation phase, let $t = T_s$. Then, (5.17) becomes

$$W_{12}^I(T_s, f) = 2|A_1 A_2| W_0\left(\frac{T_s}{2}, f\right) \cos[\Theta]. \quad (5.21)$$

Focus on the oscillation part $\cos[\Theta]$ rather than the constant part $2|A_1 A_2|$ and the envelope $W_0(T_s/2, f)$. The oscillation phase can be replaced by the coordinate of the first local maximum of $W_{12}^I(T_s, f)$. Let $\cos[\Theta] = 1$, which is equivalent to $\Theta = 0$. Then expression is obtained as:

$$f_{coordinate} = \frac{-\phi_{12}}{2\pi T_s} = \frac{-\Delta\theta}{2\pi T_s}. \quad (5.22)$$

The use of (5.22) enables one to acquire a phase shift using frequency coordinates. With respect to the modulated PSK signal, as the order increases, it will have additional kinds of phase shifts. This phase feature is extracted from PWVD.

5.3 Classifier

While most studies focused on classical clustering tools, modern clustering algorithms have an advantage over classical ones in environments that are much more hostile. To achieve clustering without a cluster number, density-based spatial clustering of applications with noise (DBSCAN) is chosen as the clustering algorithm. More importantly, compared with other complicated classifiers, such as SVM and ANNs, training, which is a very time-consuming step, is unnecessary for DBSCAN.

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

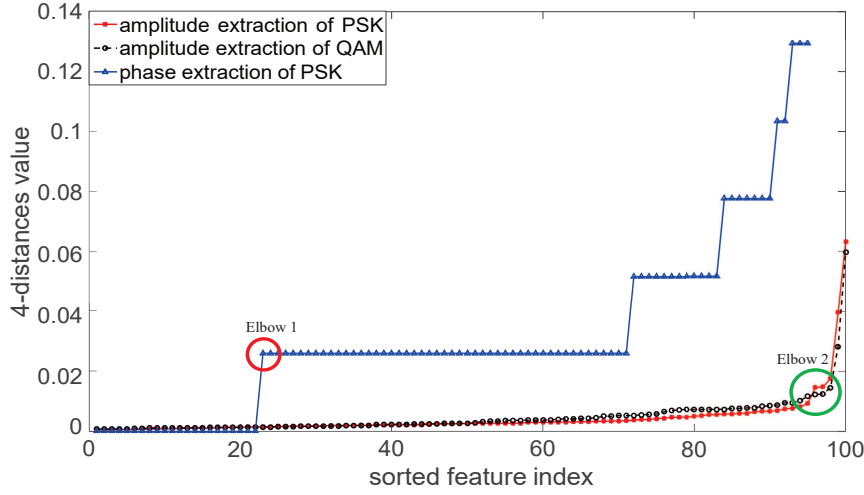


Figure 5.2: Four-distance plot of amplitude extraction and phase extraction.

5.3.1 DBSCAN

In pseudocode, the algorithm can be expressed as follows [54]:

input:

- the dataset objects.
- *Epsilon* is the radius of a neighborhood.
- *MinPts* specifies the density threshold of dense regions.

output: Density-based Clusters

Method:

mark all objects as **unvisited**;

do

 randomly select an **unvisited** object t ;

 mark t as **visited**;

if the *Epsilon* neighborhood of t has at least *MinPts* objects

 create a new cluster C , and add t to C ;

 let N be the set of objects in the *Epsilon* neighborhood of t ;

for each point t' in N

```
    if  $t'$  carries the label "unvisited"  
        DBSCAN marks it as visited;  
        if the Epsilon neighborhood of  $t'$  has at least MinPts objects  
            those objects in the Epsilon neighborhood of  $t'$  are added to  $N$ ;  
        if  $t'$  is not yet a member of any cluster, add  $t'$  to  $C$   
    end for  
    output  $C$ ;  
    else mark  $t$  as noise;  
until no object is unvisited;  
Finish
```

Epsilon and MinPts have to be determined before clustering. Epsilon is a radius that can be regarded as a minimum cluster size. It is obvious that random values offer only an unsuitable clustering result. As a heuristic, an approach called the k -distance plot is to watch the behavior at a distance from a point to its k -nearest point.

5.3.2 k -distance plot

For a given value of k , a k -distance plot goes through the distances between one point and its k th nearest point. This is a useful tool that enables one to look at the feature within clusters. Therefore, a four-distance plot shows what the distances are from every point to its 4th nearest point. Then, the result is incremented to find a sharp change, called an elbow, whose coordinates represent a suitable value of Epsilon. A small k -distance value implies groups of points that belong to several clusters. As MinPts and epsilon are coupled, as long as MinPts is set, the other one can be determined naturally.

Alternatively, MinPts is supposed to exceed data dimensions. If MinPts is too small, then noise may be incorrectly labeled as clusters. On the contrary, too large a MinPts value will improperly label the cluster as noise. In the original DBSCAN algorithm, MinPts is set as 4, which is a reasonable value for a two-dimensional (2D) data set [55]. Then, an elbow in the k -distance plot is determined, where the vertical coordinate represents epsilon. Figure 5.2 exhibits the four-distance

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

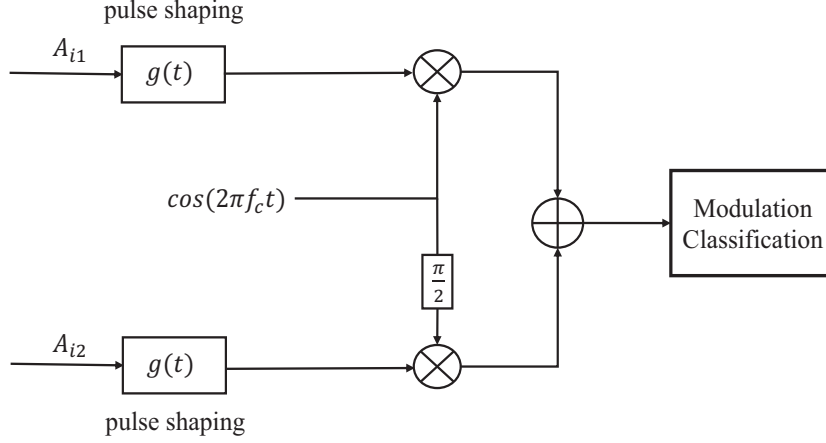


Figure 5.3: Workflow of the signal preprocessing.

plot amplitude extraction of *MPSK* and *MQAM*, together with the four-distance plot for the *MPSK* phase extraction. The red circle marks the elbow of phase extraction and the green circle marks the elbow of amplitude extraction.

5.4 Classification Scheme

In a CR system, some fundamental information about the primary user is often accessible. Moreover, because digital modulation has better immunity to noise, BPSK, QPSK, 8PSK, 16QAM, as well as 64QAM are widely used in CR systems. Therefore, they are mostly discussed in literature pertaining to modulation classification [14][56]. The workflow of this method is given by Figure 5.3.

Throughout this chapter, the assumption is established that there is a single carrier-transmitted signal, whose possible modulation type includes BPSK, QPSK, 8PSK, 16QAM, and 64QAM, and the goal is to classify the modulation type that it is using. As a hierarchical classification system, the modulation class has to be identified, after which the modulation order can be determined. Here, the characteristic that different modulation types generate different numbers of clusters is utilized.

In this paper, clustering plays the role of returning the number of clusters. Because the number of clusters corresponds to the types of modulation, the best

Table 5.1: Theoretical numbers of clusters.

	BPSK	QPSK	8PSK	16QAM	64QAM
amplitude clustering	1	1	1	3	9
phase clustering	3	5	9	—	—

way to evaluate the validation about clustering is to use the classification performance, rather than the clustering validation measures, which are most useful for deciding the k parameter of k -means [57]. Unlike k -means, DBSCAN does not assume the number of clusters. Furthermore, this research is concerned about whether the DBSCAN can offer a correct classification, and that its focus is not to minimize several quality numbers. However, to better understand the performance of this method, the sum of square error (SSE) was investigated to gain insight into the validation of DBSCAN. The theoretical number of clusters is given in Table 5.1.

16QAM and 64QAM can be classified from M PSK owing to their multi-amplitude feature. The amplitude-clustering result is identical to the symbol energy. Specifically, the in-phase and quadrature components of 16QAM take values of $\pm 1d$ and $\pm 3d$, which generates three different kinds of energy level in symbols. However, 64QAM takes values of $\pm 1d, \pm 3d, \pm 5d$, and $\pm 7d$, which generates nine different kinds of energy level in symbols. On the contrary, M PSK, which has no direct envelope information, forms only one cluster. An example of amplitude-feature clustering is illustrated in Figure 5.4, where different colors and marks represent different clusters. The SNR is 10 dB. Those clusters are recognized by DBSCAN. Although the amplitude feature that we extracted is one-dimensional (1D), Figure 5.4 represents as 2D by building a plane with the same horizontal and vertical axes; this enables more easily understanding the clustering result.

The recognition within M PSK is achieved by phase clustering. It is clear that 8PSK generates the largest number of potential oscillation phases, while BPSK

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

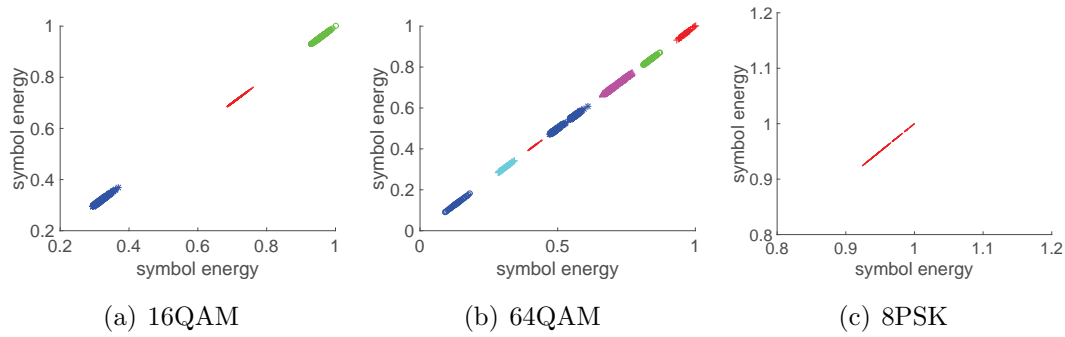


Figure 5.4: An example of amplitude-feature clustering (SNR = 10 dB).

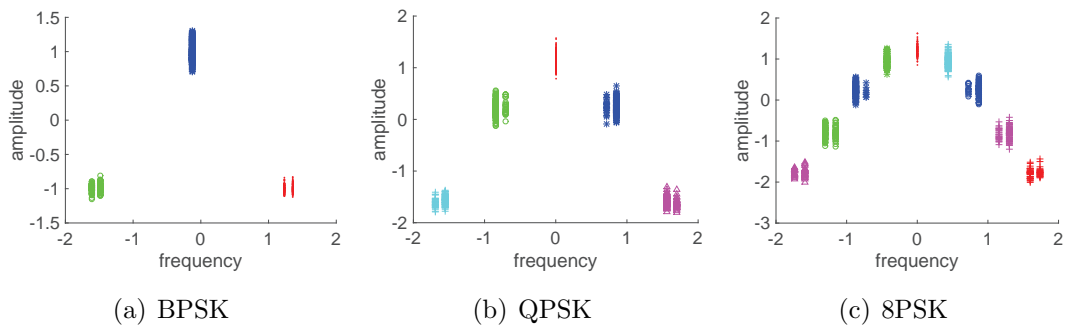


Figure 5.5: An example of phase-feature clustering (SNR = 10 dB).

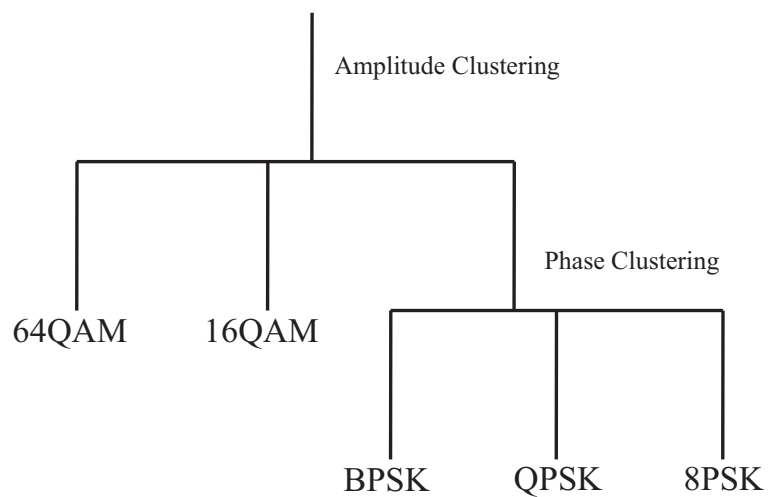


Figure 5.6: Classification workflow.

generates the fewest. More specifically, the phase shift of BPSK takes values of $[-\pi, 0, \pi]$, while the phase shift of QPSK takes values of $[-\pi, -\pi/2, 0, \pi/2, \pi]$. However, the phase shift of 8PSK takes values of $[-\pi, -3\pi/4, -\pi/2, -\pi/4, 0, \pi/4, \pi/2, -3\pi/4, \pi]$. Therefore, with respect to phase clustering, BPSK, QPSK, and 8PSK generate three clusters, five clusters, and nine clusters, respectively. An example of phase-feature clustering is given by Figure 5.5, where different colors and marks represent different clusters. The SNR is 10 dB. Those clusters are recognized by DBSCAN. The phase feature comes from the interference terms, and these interference terms have the same envelope as the signal terms. Therefore, the shapes of the plots above are determined by the Nyquist filter. The classification workflow is shown in Figure 5.6.

The amplitude clustering is first used to determine whether the modulation is amplitude-oriented (16QAM or 64QAM). The number of clusters can be used to differentiate within each subclass. If the modulation is unrelated to amplitude (MPSK), then phase clustering is used to distinguish within each subclass.

5.5 Simulation and Results

Table 5.2: Simulation parameters

carrier frequency (f_c)	sampling frequency (f_s)	symbol length	roll-off factor
$2.4 \times 10^3 Hz$	$1.2 \times 10^4 Hz$	500	1

A direct comparison cannot be made because different classifiers are designed for specific unknown parameters, such as the carrier phase and timing offset. Therefore, the performance of the proposed method is first compared with the ML and KNN in the ideal scenario, where no timing offset is involved.

All parameters are shown in Table 5.2. Figure 5.7 indicates the variation of the phase-clustering results with SNR. BPSK generates three different phase

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

shifts, which results in three or fewer clusters if the noise power increases. However, QPSK and 8PSK generate five and nine clusters, respectively. The number of clusters decreases if SNR degrades. Figure 5.8 shows the variation of the amplitude-clustering results with SNR. 16QAM generates three different symbol energy levels, which results in three or fewer clusters if there is a lot of noise. However, because of its amplitude feature, 64QAM generates nine or fewer clusters unless the SNR degrades seriously. MPSK, which has no direct envelope information, forms only one cluster.

The results of Figure 5.7 and Figure 5.8 are identical to Table 5.1 and provide an outright proof of clustering validations.

The performance of the proposed method is compared with the k-nearest neighbor [23] and ML [19] methods in Figure 5.9. For each value of the SNR, 10,000 iterations of the test data were produced. All five modulation types are considered simultaneously.

The ML shows a P_{cc} of $> 95\%$ at 8 dB, while my method shows a P_{cc} of $> 99\%$ at 8 dB. The superior performance of the proposed method is, therefore, apparent. Moreover, this method performs better than KNN, although KNN utilizes higher-order cumulants and supervised machine learning.

The influence of the signal-sampling synchronization is investigated over the individual modulations in Figure 5.10. First, it is clear that a lower SNR de-

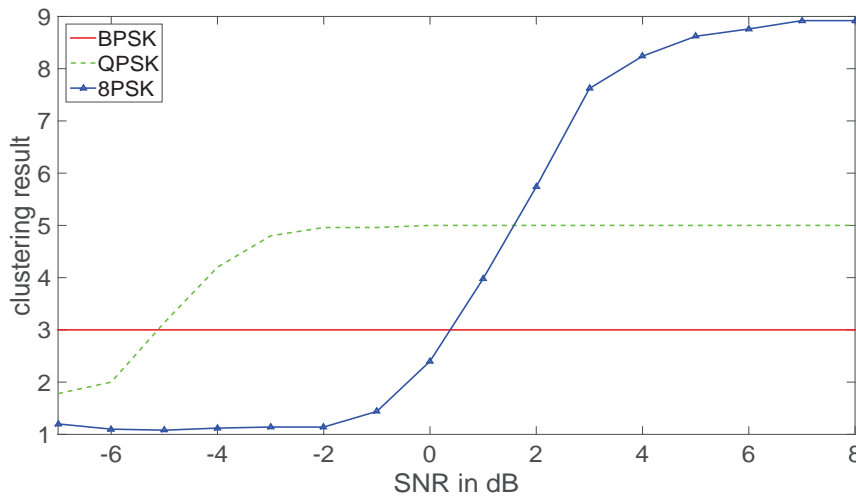


Figure 5.7: Phase-cluster number within MPSK for different SNR.

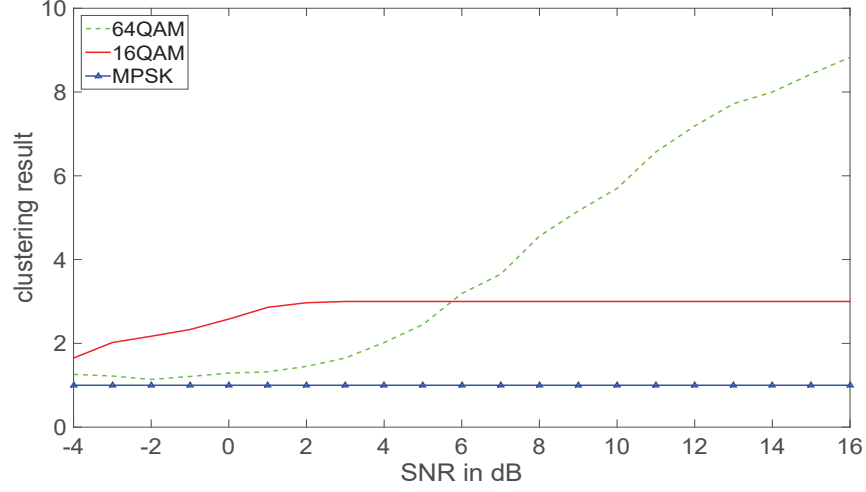


Figure 5.8: Amplitude-cluster number within *MQAM* for different SNR.

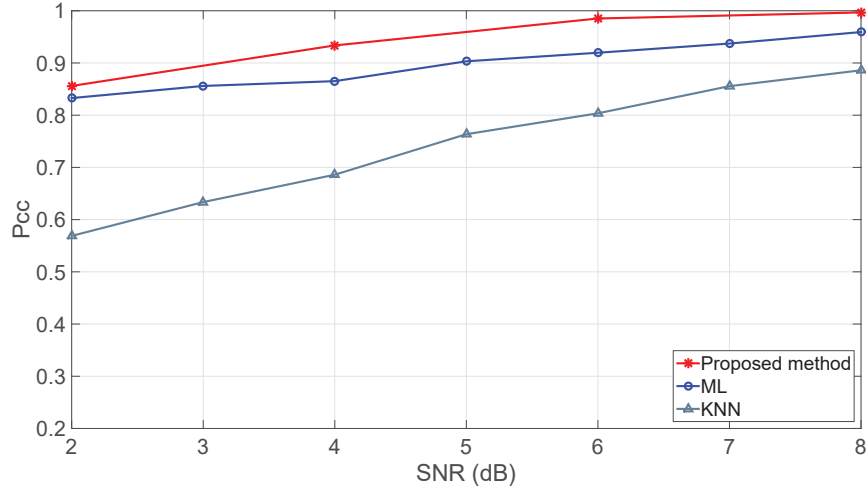


Figure 5.9: Comparison of classification accuracy for our method with ML and KNN.

grades the classification accuracy for all modulations. Of all the modulations, 64QAM has the lowest classification accuracy than others. MPSK has a higher classification accuracy than *MQAM* at lower SNRs. In addition, the raised cosine transmit filter is configured to has 20 output samples per symbol. The timing error was increased from one sample to three samples. Therefore, the timing offsets are 0.05 and 0.15, respectively. Only *MQAM* appears to be affected by

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

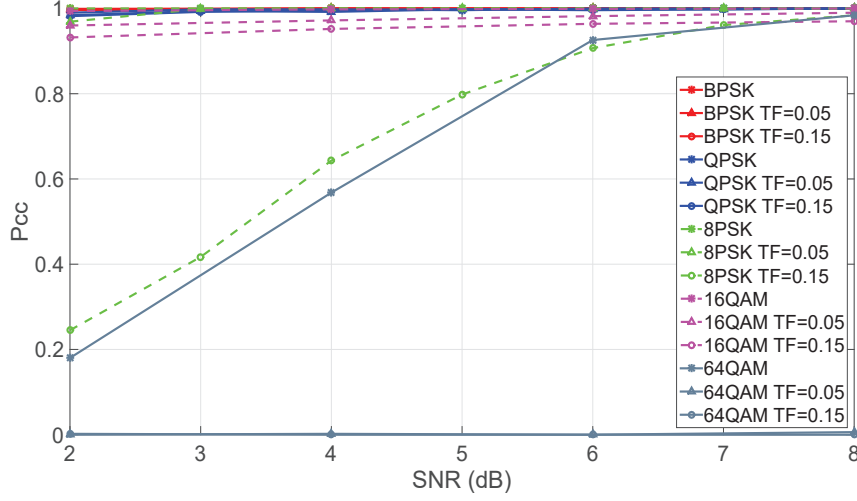


Figure 5.10: Influence of the signal-sampling synchronization on individual modulations.

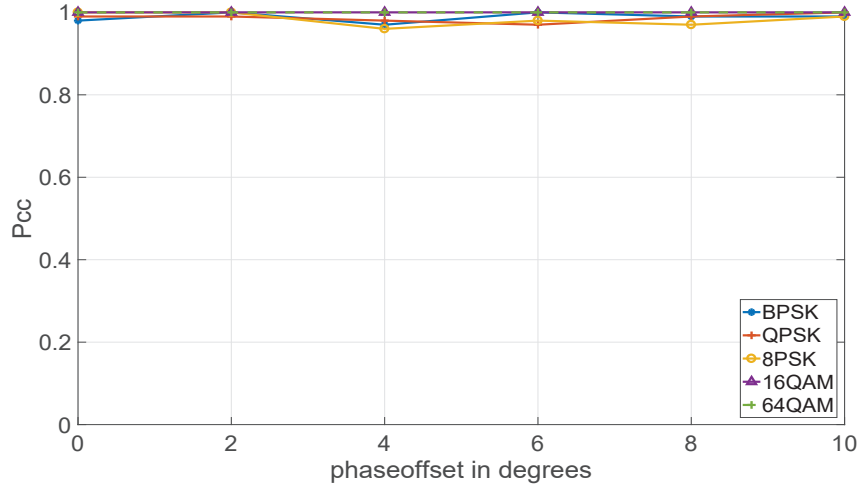


Figure 5.11: Influence of the phase offset on individual modulations.

the timing offset, and the *MPSK* shows good robustness to timing offset.

The influence of the phase offset over the classification ability is also investigated in Figure 5.11. The proposed method shows good robustness to phase offset for all five modulations. The performance of the proposed method was then compared with the KNN and ML against phase offset in Figure 5.12. It is clear that a higher phase offset degrades the classification accuracy of the ML and the

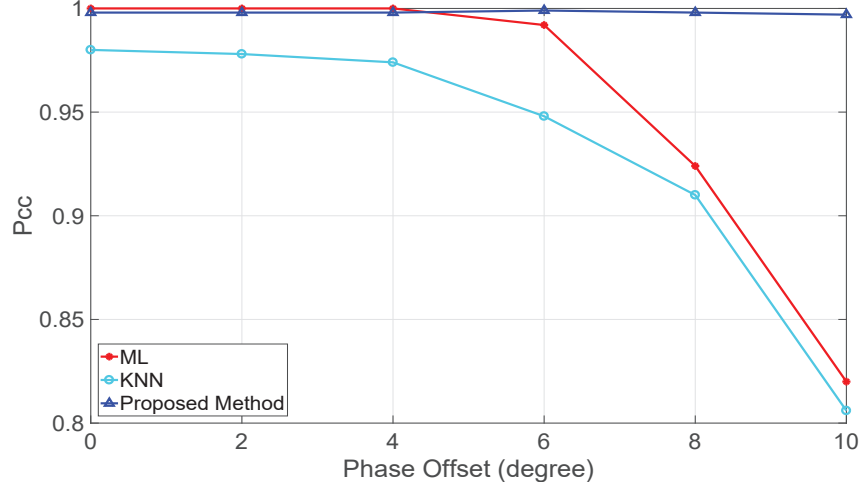


Figure 5.12: Comparison of classification accuracy of our method with ML and KNN against phase offset.

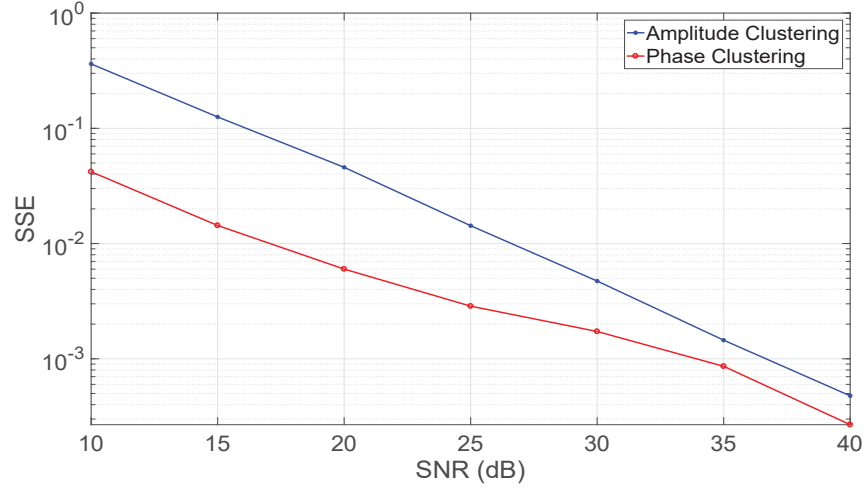


Figure 5.13: Plot of sum-of-square error vs. SNR.

KNN, which indicates that ML and KNN are extremely sensitive to phase offset.

To better understand the performance of the clustering, the SSE is used to evaluate the clustering validation in Figure 5.13, and to do this, the clustering results are compared with real classes. The higher SNR generates the lower SSE. The whole SSE does not exceed 10^{-3} if noise is absent. The trend of SSE against SNR is convincing proof of clustering validations.

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

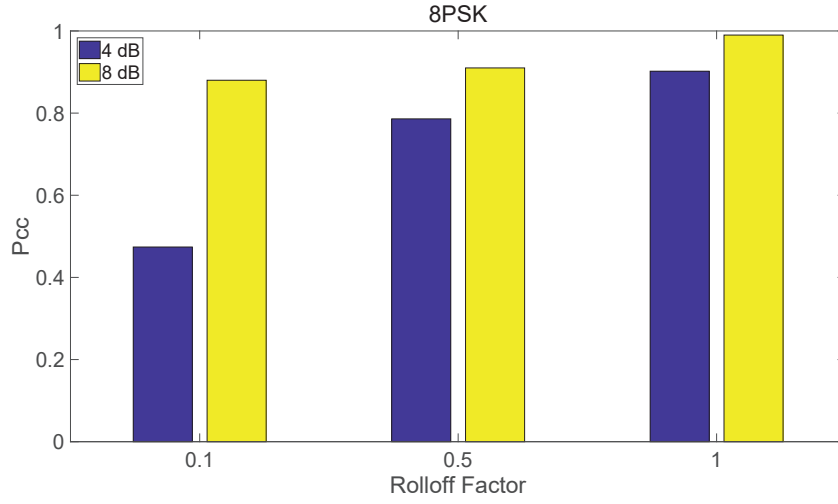


Figure 5.14: Influence of the roll-off factors on 8PSK.

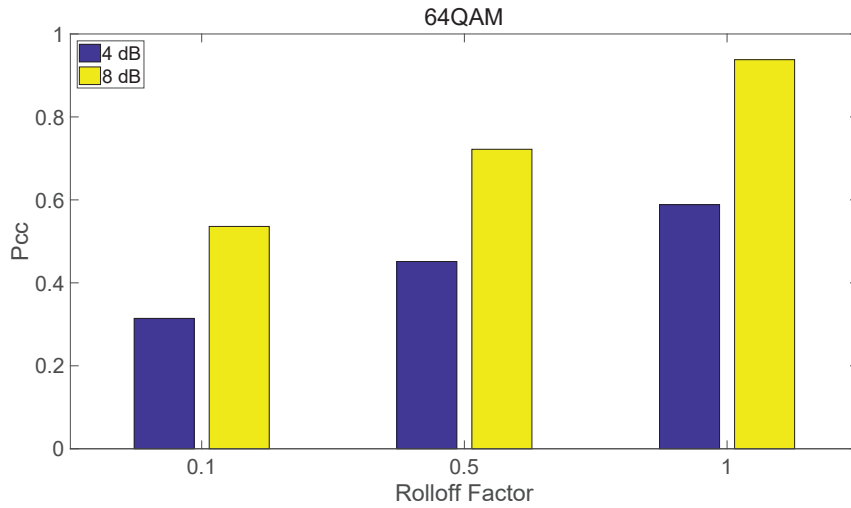


Figure 5.15: Influence of the roll-off factors on 64QAM.

Since we use shaping pulse to process the original symbols, it is necessary to investigate the influence of the roll-off factors over the classification ability. All of the above results are obtained by the raised cosine FIR pulse-shaping filter with a roll-off factor of 1. Next, the other two different roll-off factors are also taken into account. The influence of the roll-off factors on 8PSK is shown in Figure 5.14. Obviously, under the two SNRs, the higher the roll-off factor, the better the classification performance. The influence of the roll-off factors on 64QAM is

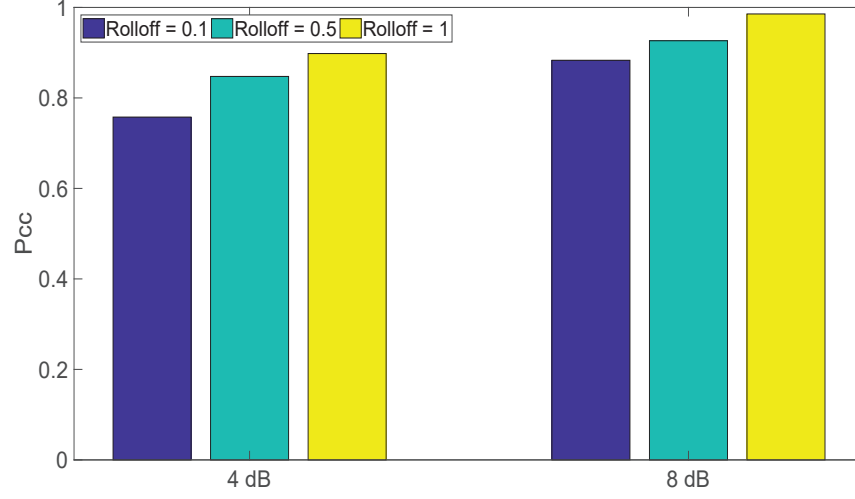


Figure 5.16: Influence of the roll-off factors on classification ability.

shown in Figure 5.15. With the rise in roll-off factor, the trend of classification performance is consistent with Figure 5.14. The influence of the roll-off factor is investigated on 5 modulations in Figure 5.16. It can be determined that the higher the roll-off factor, the higher the probability of successful classification.

5.6 Chapter Summary

Table 5.3: A summary of the three methods

	P_{cc}	Phase offset
Proposed method	High classification accuracy	Robust to phase offset
ML	Medium classification accuracy	Sensitive to phase offset
KNN	Lowest classification accuracy	Sensitive to phase offset

In this study, a classification method for specific modulation types was proposed that are used for CR. A summary of the proposed method, ML, and KNN

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

is given in Table 5.3.

The simulation results reveal that the proposed method is more accurate than the ML method because the proposed method utilized the advantages of the time-frequency features. The ML shows a P_{cc} of $> 95\%$ at 8 dB, while the proposed method shows a P_{cc} of $> 99\%$ at 8 dB. In addition, my method achieves a higher classification accuracy than the KNN in the ideal scenario. The KNN shows a P_{cc} of $> 78\%$ at 5 dB, while the proposed method shows a P_{cc} of $> 97\%$ at 5 dB. The proposed method is therefore superior to KNN in terms of classification accuracy.

To date, there is little literature that reports the influence of the signal-sampling synchronization over the classification ability, although it is a significant issue in the modulation classification field. The influence of the timing offset was investigated over the classification accuracy. In the timing offset scenario, our method shows good robustness to the timing offset for the *MPSK*. This significant characteristic makes the proposed method more practical. With respect to *MQAM*, a degradation of accuracy is foreseeable because amplitude-oriented features are sensitive to timing offsets. This shortcoming can be overcome if other expert features are used instead.

The influence of the phase was also investigated offset over the classification accuracy. My method is not sensitive to phase offset, which is proved by the Figure 5.11. Therefore, a knowledge of the carrier phase is not required for our proposed method. Actually, a phase offset rotates the constellation around the coordinate origin, which does not affect the symbol power or the phase shift between two symbols. Therefore, the phase information is unnecessary. On the contrary, the ML and KNN cannot function as usual when the phase offset increases. This is expected because ML requires all channel parameters to be known. A severe phase offset leads to a mismatch between the model used in the classifier design and the actual statistics in the channel. Therefore, one can conclude that the proposed method is superior to the ML and KNN when carrier phase is not synchronized.

Finally, the clustering validation by performing the SSE test was proved, where the higher SNR generates the low SSE, although the best way to evaluate the validation about clustering is the classification performance. The trend of SSE

against SNR is convincing proof of clustering validations. It is expected that the SSE can be further reduced if noise is excluded. In addition, according to Figure 5.7 and Figure 5.8, the number of clusters is identical to the theoretical values under high observation. This is another outright proof of clustering validation.

The drawback of this method is that to input a real signal into PWVD, the signal has to be shifted to a carrier frequency, which imposes a higher requirement for sampling. Fortunately, there are no strict requirements on the carrier frequency.

5. UNSUPERVISED MACHINE LEARNING BASED MODULATION CLASSIFICATION

Chapter 6

Conclusion

This chapter concludes the research work on the study of the supervised machine learning based modulation classifications and the unsupervised machine learning based modulation classifications in Cognitive Radio (CR) systems. The comparison between supervised classification and unsupervised classification is given first. Then, the contribution and advantages of the proposed methods are described. The future research direction is discussed at the end of this chapter.

6.1 Methods Comparisons

A comparison between supervised classification and unsupervised classification is given in Table 6.1. The ML and KNN are included as well. The symbol “—” represents that the performance has not been tested in this aspect. Only high accuracy scenario is discussed here. The first thing to emphasize is that ML is always the optimal method when the same features are used and all prior information is known. However, the ML method has not yet been proposed based on expert features.

One can see that the proposed methods by this thesis show higher classification accuracy than classical methods. On the one hand, the supervised method is the most outstanding among the others in classification ability. On the other hand, the unsupervised method provides a trade-off between easy implementation and high accuracy, since training is unnecessary for it. In addition, the supervised method has a strong resistance to timing offset, which is a superiority in

6. CONCLUSION

Table 6.1: A comparison between supervised classification and unsupervised classification

	Supervised Classification	Unsupervised Classification	ML	KNN
Required SNR (lower is better)	5 dB	7 dB	9 dB	more than 10 dB
Training	Necessary	Unnecessary	Unnecessary	Necessary
Timing Offset	Robust	Sensitive	Sensitive	Sensitive
Phase Offset	—	Robust	Sensitive	Sensitive

practicability. Regarding phase offset, the unsupervised method exhibits a strong resistance to it, which is a superiority in fading channel.

6.2 Contribution and Advantages

The main purpose of modulation classification is to provide the correct classification on the modulation of the intercepted signal from a range of modulation pools, or otherwise to decide that the modulation cannot be recognized. The above characteristics are very meaningful for cognitive radio [25][58]. In the spectrum sensing part, conventional methods, such as the energy detection, can only detect channel occupancy. If the signal is occupied by noise or other unrelated signals, the conventional method will still misjudge the presence of the primary user. This will result in a reduction in frequency utilization. However, modulation classification can distinguish between noise and artificially produced signals. It is also possible to identify the primary user-specific modulations to overcome the shortcomings of the conventional spectrum sensing techniques, due to the fact that primary users utilize one or several fixed modulations. Therefore, the

modulation detection entails the primary user detection in the cognitive radio system.

In this thesis, the modulation classification algorithms using machine learning are investigated and new methods are proposed. Firstly, a supervised machine learning based modulation classification algorithm was proposed. The higher-order cumulants as features were selected. Stacked denoising autoencoders, which is an extended edition of the neural network, was chosen as a classifier.

Secondly, an unsupervised machine learning based modulation classification algorithm was proposed. The features from time-frequency distribution were extracted. Density-based spatial clustering of applications with noise (DBSCAN) was used as a classifier because it is impossible to decide the number of clusters in advance.

6.2.1 Supervised Classification

For the supervised approach, many contributions and advantages can be seen. When two classification scenarios are considered, one of them is a rapid classification and the other one is a high accuracy scenario. For a considered rapid scenario, so far, there is few of paper that mentions this scenario, although the processing time is the priority and classification accuracy can be compromised in special conditions. Its performance evaluation shows a significant speed advantage over the conventional ML method. In addition, in a rapid classification scenario, expert features are not necessary which results in that feature extraction, which is compulsory in most conventional methods, is omitted here. It simplifies the procedure of the modulation classification and renders rapid classification more achievable. The high accuracy classification scenario is also well considered in the thesis. Although expert features are utilized as network inputs, the SDAE improves the noise resistance performance. Performance evaluation shows an accuracy advantage over the conventional ML method and the feature-based method. So far, there are few deep networks based literature that exhibits the influence of the signal sampling synchronization over the classification ability, although it is a significant issue in the modulation classification field. The influence of the timing offset over the classification accuracy are investigated in

6. CONCLUSION

both the scenarios. This ensures a comprehensive evaluation to understand the ability of this method. In the high accuracy scenario, our method is robust to the timing offset. In this research, not only a ready network structure but also the selection methods of the network structure are presented. The performance of the proposed approach is also investigated on different individual modulations. This gives the insights into the classification ability of this method.

6.2.2 Unsupervised Classification

For the unsupervised Classification method, many contributions and advantages also can be seen. The proposed method has better performance with respect to its noise resistance which is based on the proposed new features using time-frequency distribution. These features show good robustness to noise. The performance evaluation shows an accuracy advantage over the conventional maximum likelihood (ML) method and K-Nearest Neighbors (KNN) method. The novel method is robust to phase offsets, which always degrade the performance of likelihood-based methods and KNN method in the classification of PSK and QAM modulation schemes. To date, there are few reports in the literature that exhibit the influence of the signal sampling synchronization over the classification ability, although it is a significant issue in the field of modulation classification. The thesis investigated the influence of the timing offset on the classification accuracy. This ensures a comprehensive evaluation to understand the capabilities of this method. This method shows good robustness to timing offsets for *MPSK*. The proposed investigated the clustering validation against SNR and showed that DBSCAN is valid for this method. DBSCAN does not require the time-consuming training of the classifier. This is a significant advantage over supervised classifiers when rapid processing is expected. In addition, the proposed method is simple to obtain the parameters of this method. On the contrary, the cumulant-based approach needs a complicated process to determine its decision rules.

6.3 Future Research Work

Although OFDM has been widely used in modern communication technology, digital modulation technology is still an indispensable part. If the parameters of OFDM system can be detected in advance, this study can be applied directly to multi-carrier modulation system. Although there have been some scholars proposed some effective OFDM parameters detection method, to better utilize this research, the OFDM parameters detection based on machine learning are becoming increasingly significant.

6.3.1 Methods Improvement

- Supervised Classification

For the future work, algorithm optimization is still a challenging task, because the computation capability of hardware is limited. On one hand, GPU computation is a potential solution for supervised machine learning, to reduce the execution cost. Altering the modulation pool, the network structure has to be redesigned. On the other hand, that shifting the symbols to carrier frequency consumes extra computing resources. Moreover, the performance is subject to timing offset.

- Unsupervised Classification

One drawback of this method is that to input a real signal into PWVD, the signal has to be shifted to a carrier frequency, which imposes a higher requirement for sampling. If in-phase component and quadrature component can be processed separately, a real signal will be unnecessary. Moreover, DBSCAN is sensitive to parameters, which prompt us to find more robust classifiers.

6.3.2 Function Extensions

In this thesis, classifications on *MPSK* and *MQAM* have been discussed. However, Orthogonal Frequency Division Multiplexing (OFDM) is dominating digital communications, because of its robustness to interference and high spectral efficiency. Therefore, it's far more likely that next research focus would be OFDM-related classifications, e.g. OFDMs with varied pilots.

6. CONCLUSION

References

- [1] P. Kolodzy. Spectrum policy challenges. In *Invited, Interchange Conference between FCC and Academic Researchers*, 2003. 2
- [2] I. Md Habibul, K. Choo Leng, and etc. Spectrum survey in singapore: Occupancy measurements and analyses. In *Cognitive Radio Oriented Wireless Networks and Communications, 3rd International Conference on*, pages 1–7, 2008. 2
- [3] D. Datla, AM. Wyglinski, and GJ. Minden. A spectrum surveying framework for dynamic spectrum access networks. *IEEE Transactions on Vehicular Technology*, 58(8):4158–4168, 2009. 2
- [4] J. Mitola and GQ. Maguire. Cognitive radio: making software radios more personal. *IEEE personal communications*, 6(4):13–18, 1999. 2
- [5] J. Mitola. *Cognitive radio-an integrated agent architecture for software defined radio*. Royal Institute of Technology (KTH), 2000. 2
- [6] S. Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE journal on selected areas in communications*, 23(2):201–220, 2005. 2
- [7] IF. Akyildiz, WY. Lee, and KR. Chowdhury. Crahns: Cognitive radio ad hoc networks. *AD hoc networks*, 7(5):810–836, 2009. 3, 4
- [8] IF. Akyildiz and WY. Lee etc. Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer networks*, 50(13):2127–2159, 2006. 4
- [9] T. Yucek and H. Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE communications surveys tutorials*, 11(1):116–130, 2009. 5
- [10] J. Ma, G. Li, and BH. Juang. Signal processing in cognitive radio. *Proceedings of the IEEE*, 97(5):805–823, 2009. 5
- [11] S. Haykin, DJ. Thomson, and JH. Reed. Spectrum sensing for cognitive radio. *Proceedings of the IEEE*, 97(5):849–877, 2009. 5

REFERENCES

- [12] Y. Zeng, Y. Liang, and A. Hoang etc. A review on spectrum sensing for cognitive radio: challenges and solutions. *EURASIP Journal on Advances in Signal Processing*, 2010(1):381465, 2010. 5
- [13] Q. Shi and Y. Karasawa. Automatic modulation identification based on the probability density function of signal phase. *IEEE Transactions on Communications*, 60(4):1033–1044, 2012. 6
- [14] M. Abdelbar, B. Tranter, and T. Bose. Cooperative modulation classification of multiple signals in cognitive radio networks. In *IEEE International Conference on Communications*, pages 1483–1488, Sydney, NSW, 2014. 7, 72
- [15] X. Zhu and T. Fujii. Modulation classification in cognitive radios for satellite and terrestrial systems. In *IEEE International Conference on Communications Workshop*, pages 1612–1616, London, UK, 2015. 7
- [16] N. Lay and A. Polydoros. Modulation classification of signals in unknown isi environments. In *IEEE Conference on Military Communications*, pages 170–174, San Diego, CA, 1995. 7
- [17] L. Hong and KC. Ho. An antenna array likelihood modulation classifier for bpsk and qpsk signals. In *IEEE Conference on Military Communications*, pages 647–651, Anaheim, CA, 2002. 7, 22, 23
- [18] A. Abdi, OA. Dobre, R. Choudhry, Y. Bar-Ness, and W. Su. Modulation classification in fading channels using antenna arrays. In *IEEE Conference on Military Communications*, pages 211–217, Monterey, CA, 2004. 7
- [19] W. Wei and J.M. Mendel. Maximum-likelihood classification for digital amplitude-phase modulations. *IEEE Transactions on Communications*, 48(2):189–193, 2000. 7, 10, 50, 57, 76
- [20] A. Swami and BM. Sadler. Hierarchical digital modulation classification using cumulants. *IEEE Transactions on Communications*, 48(3):416–429, 2000. 8, 10
- [21] CM. Spooner. On the utility of sixth-order cyclic cumulants for rf signal classification. In *the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, pages 890–897, Pacific Grove, CA, 2001. 8
- [22] M. Bkassiny, Y. Li, and SK. Jayaweera. A survey on machine-learning techniques in cognitive radios. *IEEE Communications Surveys and Tutorials*, 15(3):1136–1159, 2012. 8
- [23] MW. Aslam and AK. Nandi. Automatic modulation classification using combination of genetic program and knn. *IEEE Transactions on Wireless Communications*, 11(8):2742–2750, 2012. 10, 57, 76

REFERENCES

- [24] T. J O'Shea, J. Corgan, and T. C. Clancy. Convolutional radio modulation recognition networks. In *17th International Conference on Engineering Applications of Neural Networks*, pages 213–226, Aberdeen, UK, 2016. 11, 18, 40, 61
- [25] Z. Zhu and A.K. Nandi. *Automatic modulation classification: principles, algorithms and applications*. Wiley, United Kingdom, 2015. 17, 23, 86
- [26] B. Migliori, R. Zeller-Townson, D. Grady, and D. Gebhardt. Biologically inspired radio signal feature extraction with sparse denoising autoencoders. *arXiv preprint arXiv:1605.05239*, 2016. 18, 40
- [27] X. Zhu and T. Fujii. Modulation classification for cognitive radios using stacked denoising autoencoders. *International Journal of Satellite Communications and Networking*, 2016. 24
- [28] EE. Azzouz and AK. Nandi. Procedure for automatic recognition of analogue and digital modulations. *IEE Proceedings Communications*, 143(5):259–266, 1996. 25, 26, 39
- [29] Y. Chan, L. Gadbois, and P. Yansouni. Identification of the modulation type of a signal. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP*, volume 10, pages 838–841, Apr 1985. 26
- [30] S. Soliman and S. Z. Hsue. Signal classification using statistical moments. *IEEE Transactions on Communications*, 40(5):908–916, May 1992. 29, 42
- [31] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. 35
- [32] NS. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992. 35
- [33] H. Agirman-Tosun, Y. Liu, and etc. Modulation classification of mimo-ofdm signals by independent component analysis and support vector machines. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1903–1907, 2011. 36
- [34] M.S. Mhlhaus, M. ner, and etc. Automatic modulation classification for mimo systems using fourth-order cumulants. In *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5, 2012. 36
- [35] K. Haq, A. Mansour, and S. Nordholm. Recognition of digital modulated signals based on statistical parameters. In *IEEE International Conference on Digital Ecosystems and Technologies*, pages 565–570, Dubai, UAE, 2010. 39, 64

REFERENCES

- [36] P. Vincent, H. Larochelle, Y. Bengio, and PA. Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1096–1103, Helsinki, Finland, 2008. 40, 47
- [37] GE. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006. 40
- [38] B. M. Sadler, G. B. Giannakis, and Keh-Shin Lii. Estimation and detection in non-gaussian noise using higher order statistics. *IEEE Transactions on Signal Processing*, 42(10):2729–2741, 1994. 41, 57
- [39] WA. Gardner and CM. Spooner. The cumulant theory of cyclostationary time-series. i. foundation. *IEEE Transactions on Signal Processing*, 42(12):3387–3408, 1994. 42
- [40] Bengio, Yoshua, Lamblin, Pascal, Popovici, et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007. 44
- [41] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2094–2107, 2014. 45
- [42] Bengio and Yoshua. Learning deep architectures for ai. *Foundations and trends*. 45, 57
- [43] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and PA. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, (11):3371–3408, 2010. 45
- [44] L. Bottou. Large-scale machine learning with stochastic gradient descent. In *19th International Conference on Computational Statistics*, pages 177–186, Paris, France, 2010. 48
- [45] D. R. Pauluzzi and N. C. Beaulieu. A comparison of snr estimation techniques for the awgn channel. *IEEE Transactions on Communications*, 48(10):1681–1691, 2000. 54
- [46] I. T. Nabney. Efficient training of rbf networks for classification. In *Ninth International Conference on Artificial Neural Networks*, pages 210–215, Edinburgh, UK, 1999. 57
- [47] Z. Liu and K Yi. Symbol timing synchronization using interpolation-based matched-filters. *Wireless Personal Communications*, 50(4):457–467, 2009. 60
- [48] S. K. Sharma, S. Chatzinotas, and B. Ottersten. Eigenvalue-based sensing and snr estimation for cognitive radio in presence of noise correlation. *IEEE Transactions on Vehicular Technology*, 62(8):3671–3684, 2013. 60

REFERENCES

- [49] H. Ketterer, F. Jondral, and A.H. Costa. Classification of modulation modes using time-frequency methods. In *IEEE International Conference on Acoustics, Speech, and Signal*, pages 2417–2474, Phoenix, USA, 1999. 63
- [50] D.Le Guen and A. Mansour. Automatic recognition algorithm for digitally modulated signals based on statistical approach in time-frequency domain. In *Baiona workshop on signal processing in communications*, pages 247–252, Spain, 2003. 64
- [51] W. Mecklenbraeuker and F. Hlawatsch. *The Wigner Distribution: Theory and Applications in Signal Processing*. Elsevier Science, Amsterdam, 1997. 64, 66, 68
- [52] E. Sejdic, I. Djurovic, and J. Jiang. Time-frequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing*, 19(1):153–183, 2009. 64
- [53] L. Cohen. *The Wigner Distribution: Theory and Applications in Signal Processing*. Prentice Hall, Upper Saddle River, 1997. 64
- [54] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco, 2011. 70
- [55] P. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison Wesley, Boston, 2005. 71
- [56] Y. Zhang, G. Memik, and J. Grosspietsch. Digital modulation classification using temporal waveform features for cognitive radios. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC2007)*, pages 1–5, Athens, Greece, 2007. 72
- [57] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu. Understanding of internal clustering validation measures. In *IEEE International Conference on Data Mining*, pages 911–916, Sydney, Australia, 2010. 73
- [58] M.R. Bahloul, M.Z. Yusoff, AH. Abdel-aty, and L. Kudis. Spectrum sensing for cognitive radio systems using modulation classification: A review. *Applied Mathematics Information Sciences*, 11(3):857–865, 2017. 86

REFERENCES

Publications

List of Publications Directly Related to Thesis

Journal Papers

1. Xu Zhu, Takeo Fujii, "Modulation Classification for Cognitive Radios using Stacked Denoising Autoencoders", International Journal of Satellite Communications and Networking, Wiley, 2016.

International Conference Papers

1. Xu Zhu, Takeo Fujii, "A Novel Modulation Classification Method in Cognitive Radios based on Features Clustering of Time-frequency", in Proceedings of *IEEE Radio and Wireless Symposium (RWS2016)*, Austin, USA, Jan. 2016.

List of Publications for References

Journal Paper

1. X.K. Li, Zhi Xia, and Xu Zhu, "Image Morphological Characteristics of Geometrical Structure of Target Echo Time-frequency Distribution", *Acta Armamentarii*, vol. 36, no. 1, pp.130-137, 2015.

PUBLICATIONS

International Conference Papers

1. Xu Zhu, Takeo Fujii, "A Modulation Classification Method in Cognitive Radios System using Stacked Denoising Sparse Autoencoder", in Proceedings of *IEEE Radio and Wireless Symposium (RWS2017)*, Phoenix, USA, Jan. 2017.
2. Xu Zhu, Takeo Fujii, "A Novel Modulation Classification Method in Cognitive Radios using Higher-Order Cumulants and Denoising Stacked Sparse Autoencoder", in Proceedings of *Asia-Pacific Signal and Information Processing Association (APSIPA2016)*, Jeju, Korea, Dec. 2016.
3. Xu Zhu, Takeo Fujii, "Modulation Classification in Cognitive Radios for Satellite and Terrestrial Systems", in Proceedings of *IEEE International Conference on Communications (ICC2015)*, London, United Kingdom, Jun. 2015.
4. Xu Zhu, Takeo Fujii, "PSK Classification based on Time-Frequency Distribution", in Proceedings of *Triangle Symposium on Advanced ICT (TriSAI2014)*, Beijing, China, Sep. 2014.
5. Xu Zhu, Takeo Fujii, "Wideband Spectrum Sensing based on Wavelet Transform", in Proceedings of *Triangle Symposium on Advanced ICT (TriSAI2012)*, Tokyo, Japan, Sep. 2012.

Domestic Conference Papers

1. Hang Liu, Xu Zhu, and Takeo Fujii, "Primary user detection in cognitive radio using spectral-correlation features and stacked denoising autoencoders based on signal classification", IEICE Technical conference on Software Radio (IEICE SR 2017), Tokyo, Japan, May. 2017.
2. Xu Zhu, and Takeo Fujii, "A Novel Modulation Classification Method in Cognitive Radios using Deep Network", IEICE Technical conference on Software Radio (IEICE SR 2016), Nagoya, Japan, July. 2016.

3. Xu Zhu, and Takeo Fujii, ”**Time-Frequency Analysis based PSK Modulation Classification**”, IEICE Technical conference on Softwared Radio (IEICE SR 2014), Kyoto, Japan, July. 2014.

PUBLICATIONS
