The Overview of Modulation Recognition of Communication Signals

Yibin Pan *

Department of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou, China

* Corresponding Author Email: dxjalyp@hotmail.com

Abstract. The modulation recognition technology assumes a key and central role in precisely identifying the modulation scheme of received signals, a task of utmost significance for various essential undertakings such as spectrum management, interference mitigation, and secure communication. The paper studies two principal approaches: model-driven and data-driven methods. Model-driven methods, exemplified by maximum likelihood estimation (MLE) and matched filtering, hinge upon theoretical models and deterministic signal characteristics. These methods possess high interpretability and exhibit efficiency in low-noise settings but are limited in adaptability to high-noise, dynamic, or nonlinear signal conditions. Data-driven methods, which capitalize on machine learning and deep learning, are capable of automatically extracting features and patterns from data. Techniques like convolutional neural networks (CNN) and transformer architectures manifest superior performance in dealing with high-dimensional, complex, and noisy signals, albeit they necessitate copious amounts of labeled data and substantial computational resources. Finally, several future research directions are put forward, such as hybrid model integration to harness the merits of both approaches, adaptive learning for dynamic environments, and optimization for real-time and resource-constrained applications, thereby laying a solid foundation and paving the path for the development of more robust and intelligent modulation recognition systems.

Keywords: Communication signal modulation recognition, model-driven, data-driven.

1. Introduction

In the rapidly evolving landscape of modern communication technology, communication signal modulation recognition has emerged as a topic of paramount importance. Communication signal modulation recognition refers to the process of accurately discerning the modulation scheme employed by the received communication signal. Serving as the linchpin between signal detection and demodulation, the primary mission of modulation recognition is to meticulously identify the modulation system of the signal through sophisticated analytical technique [1]. Accurate recognition of the modulation technique directly determines the quality of the signal in the ensuing demodulation phase.

The application scenarios of modulation recognition are extensive and diverse. In the military realm, the effective modulation recognition method allows the armed forces to intercept and dissect enemy communication signals with precision. This capability is not only instrumental in gathering intelligence but also in devising effective counter-strategies. For instance, during a military operation, by promptly recognizing the modulation type of an adversary's communication signal, the military can take immediate measures to jam or disrupt the enemy's communication channels, thereby gaining a significant tactical advantage. In the civilian sphere, modulation recognition is the cornerstone of efficient spectrum management. With the exponential growth of wireless communication devices and services, the radio frequency spectrum has become an increasingly valuable resource. Modulation recognition techniques enable regulatory bodies and service providers to monitor and manage the spectrum usage effectively. By accurately identifying the modulation schemes of various signals, they can ensure that different communication systems coexist harmoniously, minimizing interference and maximizing the utilization of the available spectrum. This, in turn, translates to enhanced quality and reliability of communication services for end-users, be it in mobile telephony, satellite communication, or wireless data transfer.

Nowadays, modulation recognition algorithms for communication signals have been extensively studied and can be broadly bifurcated into two paradigms: model-driven and data-driven. Model-driven methods formulate recognition algorithms by established theoretical models, which require an in-depth comprehension of the mathematical underpinnings and principles of communication signals. The model-driven method assumes the signal behavior can be mathematically modeled and predicted. For example, certain model-driven techniques might rely on the statistical properties of the signal, such as its mean, variance, and higher-order moments, to infer the modulation type. Conversely, data-driven methods extract meaningful features and patterns from the data itself instead of relying on pre-defined mathematical models. With the advent of big data and the continuous improvement of machine learning algorithms, data-driven methods have gained significant traction. They are particularly adept at handling complex and non-linear relationships in the data, which are often challenging to model traditional analytical methods.

This paper is structured in a systematic manner. Firstly, introduce the basic definition of the model-driven approach in communication signal modulation recognition, followed by an in-depth analysis of several archetypal methods. The strengths and weaknesses of each method will be meticulously evaluated, clarifying their suitability and limitations in different application contexts. Secondly, the data-driven approach will be subjected to a similar comprehensive scrutiny. Finally, a comprehensive summary of the full text is conducted, and the future development trajectory is predicted.

2. Model-Driven Communication Signal Modulation Status

2.1. Basic Definition

The model-based communication signal modulation identification method is realized through the theoretical knowledge and mathematical modeling means of domain experts. The core of this approach is to use the known communication signal characteristics and their mathematical description, combining personal experience and expertise, to build models that can identify the type of signal modulation. In this method, the establishment of feature extraction and identification model is completed by manual design, which fully embodies the certainty and interpretability in the communication theory [2, 3].

Among them, model-based methods are usually regarded as white-box algorithms, and their computational process is transparent and has strong interpretability [4]. Each step of this approach can be tracked and validated, and analysts are able to clearly understand the decision logic of the model and the basis for identifying the results. Therefore, it performs well in scenarios requiring high verifiability and high reliability. However, the model-based approach also has its limitations, which is highly dependent on manually extracted signal features and theoretical models and difficult to adapt to high noisy and nonlinear complex environments. Moreover, this approach often exhibits maladaptive in processing dynamically changing signals and is vulnerable to environmental uncertainty.

In conclusion, the model-based modulation recognition method has the advantages of solid theoretical foundation, interpretable results and high computational efficiency, and is suitable for the communication signal environment with low noise, obvious features and determined structure. However, in complex dynamic signal scenarios, its adaptability and robustness are weak, and it needs to be weighed according to specific application scenarios.

2.2. Comparative Analysis of Model-Based Recognition Methods

In the model-based communication signal modulation identification method, there are many classical algorithms, which have their own characteristics in the theoretical basis, applicable scenarios and performance [5]. Among them, the most representative methods include the maximum likelihood estimation (MLE) and the matching filtering method.

MLE is a statistical theory-based parameter estimation method, whose core idea is to determine the best type of modulation that best matches the received signal by maximizing the likelihood function [6]. The MLE method provides extremely high recognition accuracy under ideal conditions, especially in environments with low interference and high signal-to-noise ratio. However, this method requires more computational resources, especially when processing high-dimensional, large-scale data, and the computational complexity will increase significantly [7]. Moreover, in scenarios with high noise or severe interference, the estimation results of MLE are easy to deviate from the real situation, leading to a decrease in identification accuracy.

Matching filter method is another classical model-based method, whose basic principle is to build the pulse response filter for the target signal to enhance the received signal to improve the detection ability of the signal and suppress the noise interference [8]. Matched filtering method has the advantage of simple computation and strong real-time, especially suitable in communication systems with low signal to noise ratio [9]. However, the limitation of the method is the need for prior structural information of the target signal, which makes it underperform when processing unknown signals or a widely variable signal environment. Moreover, the matching filter method is highly sensitive to the noise, and the system performance may be significantly affected when the noise distribution is unstable.

By comparing these two methods, it can be found that the maximum likelihood estimation has the advantage of high precision in high SNR and clear signal characteristics, but its calculation complexity is high for scenarios with sufficient computing resources, while the matching filter method has strong real-time processing power under low signal / noise ratio, but it depends on prior knowledge of target signal structure, so it is difficult to deal with complex dynamic signal environment.

Overall, the model-based approach performs well in theoretical transparency and interpretability, and is suitable for well-characterized, structurally stable communication signal recognition tasks. However, different methods have their own advantages and disadvantages in performance, and the specific selection needs to comprehensively consider the signal characteristics, computational resources and application requirements. With the increasingly complex communication environment, how to further improve the adaptability of model-based methods in high-noise and dynamic environments will be an important direction in the future research.

3. Data-Driven Communication Signal Modulation Recognition Status

3.1. Basic Definition

Data-driven communication signal modulation recognition methods rely on machine learning and deep learning techniques to automatically extract signal features from large amounts of data and capture signal patterns and characteristics. In the identification process, the data-driven method will extract multidimensional features such as time domain, frequency domain and statistical characteristics, and train the identification model through the algorithm, so as to realize the classification of signal modulation mode. This process does not rely on pre-set mathematical models. Instead, it learns non-linear relationships through large-scale data to enhance the processing power of complex signal [10].

Unlike traditional identification strategies that rely on manual feature selection, data-driven methods improve the ability of the system to adapt to complex and high-noise environments through automated learning [11]. Although the data-driven method shows strong adaptability and recognition accuracy, its limitations cannot be ignored. First, data-driven methods are highly dependent on large amounts of annotated data, and model performance will be affected when the data is insufficient. Second, the method requires large computational resources, with deployment challenges for resource-constrained systems. Finally, the learning process within the data-driven model is more complex, and it is difficult to explain the decision logic of each step, thus affecting the transparency and credibility of the model.

With its automatic feature extraction and learning ability, the data-driven method performs superior in processing high-dimensional, non-linear and complex noise environment, and has become an important direction of communication signal modulation recognition research.

3.2. Comparative Analysis of Modulation Recognition Methods

Recently, in the field of communication signal modulation and recognition, the adaptable and efficient data-driven method has gained a lot of attention. In particular, three methods based on traditional machine learning, convolutional neural network (CNN) and Transformer architecture are particularly remarkable. Each of them has its own uniqueness in theoretical roots, advantages and disadvantages and application scope. This part of the article will focus on the specific comparison and analysis of these aspects, and focus on the theoretical foundation of each technology, advantages and disadvantages and the application situation that can best play its utility.

Artificial feature extraction is crucial for traditional machine learning algorithms to identify key features of the signal such as amplitude, phase and frequency domain properties. These key features are sent into classification algorithms such as support vector machines (SVM), decision trees, or k-nearest neighbors for subsequent processing.

SVM performs well in the classification task of high-dimensional feature space. It is particularly effective in dealing with small sample data and identifying complex boundaries, achieving high-precision classification by finding optimal hyperplanes between different categories [12, 13]. However, the performance of the SVM is highly dependent on the choice of kernel functions and the parameter configuration. Moreover, as the size of the dataset increases significantly, limiting its practical application in large-scale tasks.

Convolutional neural network (CNN) in the field of deep learning effectively breaks through the limitations of conventional technology, especially with the help of its unique convolution operation function, it automatically identifies the hierarchical features in signal data to replace the manual extraction needed in the past. This makes it more suitable for capturing complex properties in the time and frequency domain ranges and shows great advantages in processing data with nonlinearity or contaminated by noise. For example, traditional strategies are often weak when the SNR field is low, and then the CNN is deployed on a large scale [14]. Moreover, because of the powerful compatibility of different modulation modes, CNN becomes one of the indispensable powerful tools in communication systems nowadays. However, such technology is not perfect; the first problem is the high demand for rich computing resources and the demand of a large number of labeled training samples. Especially when the task in this scenario involves tedious data collection and annotation, the high dependence on the annotated data set may be the key link of efficiency; then the challenge to the enforceability in the real-time or resource-limited environment causes the consumption of CPU /GPU to be an important constraint.

The Transformer model shows a high-level scheme in the field of data-driven modulation recognition. This approach was originally designed to implement natural language processing tasks by using self-attention mechanisms to grasp long-term dependencies in serialized data. This outstanding feature makes the Transformer model outstanding when analyzing multidimensional signals and complex time series information. For example, in a rapidly changing dynamic communication environment, it can capture the global context and adapt to a variety of complex patterns with superior accuracy. At the same time, thanks to its parallel processing power, this architecture can be extremely efficient in large-scale data set applications. In comparison the traditional sequential algorithm, it is capable of reducing the time consumption in the training stage. However, it is worth of note that the complexity of the Transformer structure itself also poses a series of challenges. Its huge demand for computing resources and the extensive support of annotated data supports also limit its application in poor resources. And just like CNN, the "black box" positioning often makes it face questions of interpretability and easy to cause some confusion and trouble in the game, especially need to pay attention.

The results of these method comparative analyses clearly indicate their differences in their respective strengths, limitations, and best application scenarios. Classical machine learning techniques are especially suitable for handling simple tasks due to their high efficiency and simple implementation process. However, their demand for artificial feature engineering and the limitation of scalability makes such models are limited in the face of complex dynamic signals. At the same time, convolutional neural network (CNN), with its automatic feature extraction ability and powerful anti-noise interference ability, is unique in dealing with signal processing tasks with rich nonlinear features and high dimensional complexity, especially in applications that need high precision and highly adaptability. The Transformer structure, with its excellent long-term dependency capture ability and complex mode modeling performance, is very suitable for information transmission problems in challenging and rapidly changing environments. However, it is undeniable that the high computational demand and the dependence on a wide range of data sets also bring considerable obstacles and difficulties to deploy this technology in the actual system.

In general, data-driven technology plays a key role in the field of communication signal modulation and recognition. For basic and restricted application environments, traditional machine learning provides economical and efficient solutions; CNN balances accuracy and flexibility and performs well in medium complexity, and Transformer achieves the ultimate standard for performance improvement for rapidly changing and complex signal environments. Practitioners should judge the most appropriate technology choice according to the actual needs of the application, and consider the computing resources, available data and the complexity of the signal as factors to achieve the best results. In addition, integrating these methods into the new technical framework of transfer learning and joint learning will provide new ways to cope with the existing difficulties and open up new ways to adjust the recognition ability.

4. Future Development Trends

Communication signal modulation and recognition technology is making continuous progress with communication technology, especially in 5G and future 6G networks, the diversity of modulation modes and the complexity of environment put forward higher requirements for modulation recognition. In the future, the development trend of modulation recognition technology will be mainly reflected in the following aspects.

With the progress of technology, the modulation recognition approach of deep learning has become the dominant trend. In particular, the convolutional neural network CNN and Transformer model can independently grasp the core features from the massive data due to the excellent feature extraction strength, and reduce the requirements of manual design, and show its advantages under the conditions of complex signal and noisy noise. Looking ahead, more and more studies may explore more profound network structures, such as deep convolution or multi-dimensional scaling conversion structures, which may also be further expanded to improve accuracy and robust power.

In addition, the rise of hybrid models also marks a key trend in the future development of modulation recognition technology. Future explorations may focus on integrating deep learning with traditional machine learning strategies to double-strengthen their respective strengths. Suppose that traditional machine learning methods can perform excellent training efficiency when dealing with small data sets. In contrast, deep learning methods rely more on their inherent capabilities to automatically discover characteristic signs of higher complexity. The future modulation awareness structure of these two systems is expected to have greater flexibility and efficacy than today, and to provide excellent performance in diverse signals and hardware devices.

In the future development trend of modulation identification technology, data privacy and security are undoubtedly a key issue. Facing the widespread use of massive data sets, how to guarantee information privacy and realize effective modulation recognition has become one of the key problems in research. Innovations such as edge computing and federated learning have the potential to provide useful answers to this dilemma, which can keep data processing locally and further facilitate the learning process of machines, reducing the risk of personal and sensitive information leakage over long distances.

5. Conclusion

In communication signal modulation recognition, model-based and data-based methods have advantages and limitations. The model-based method relies on theoretical signal models and statistical analysis, which is characterized by simple calculation, easy operation and high interpretability, and exhibits stability in low noise and deterministic environments. However, this method is limited by prior knowledge and a fixed mathematical framework, which is difficult to adapt to complex dynamic environments, and is prone to encounter bottlenecks especially in scenes with obvious high-frequency nonlinear features. In contrast, the data-based method uses machine learning and deep learning algorithms to automatically extract features and learn patterns. It performs well in high-dimensional and complex data environment, showing strong adaptability and robustness, which is especially suitable for application in the case of high noise and diverse modulation types. However, data-driven methods rely on large-scale annotated data and high computational resources, and the effect will be significantly affected when the data is insufficient. Future modulation recognition field will be with the fusion of deep learning and traditional methods, the efficiency of the model, interdisciplinary applicability and the development of data privacy protection technology forward, promote modulation recognition technology to the direction of more intelligent, more robust and more real-time, for wireless communication system to provide more efficient and more reliable support.

References

[1] Pi S, Zhang S, Wang S, Guo B, Yan W. Improving Modulation Recognition Using Time Series Data Augmentation via a Spatiotemporal Multi-Channel Framework. Electronics, 2023, 12 (1): 96.

[2] Wang Yuhong, Fei Gaolei, Hu Guangmin. Improved maximum-likelihood-based fast topology estimation method. Computer application, 2011, 31 (1): 212 - 214, 238.

[3] Wang Hongwu, Yang Teng, Zhang Jiwei, et al. UWB channel fusion estimation method based on tree wavelet compression and maximum likelihood. Journal of Liaoning Engineering and Technical University (Natural Science Edition), 2024, 43 (3): 373 - 378.

[4] Zhao Jun. The Development and Application of Support Vector Machine. Journal of Physics Conference Series, 2021, 1748 (5): 052006.

[5] Atik, H., Kut, D., Soylu, A. Support Vector Machine Chains with a Novel Tournament Voting. Electronics, 2023, 12 (11): 2485.

[6] Cervantes, A., García, L., Rodríguez, E. A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges, and Trends. Neurocomputing, 2020, 408: 189 - 215.

[7] Pisner, D., Schnyer, D. Support Vector Machine. Machine Learning, 2020.

[8] Nandi, S., Azzouz, E. Algorithms for automatic modulation recognition of communication signals. IEEE Transactions on Communications, 2006, 54 (6): 1164 – 1170.

[9] Khan, M. S., Khan, I. Automatic modulation recognition of communication signals. International Journal of Communication Systems, 2017, 30 (6): e3094.

[10] Jiang, Z., Liu, Y., Li, W. Modulation recognition of communication signal based on convolutional neural network. IEEE Transactions on Signal Processing, 2021, 69: 1095 - 1106.

[11] Li, Y., Shi, X. Modulation recognition algorithm based on deep learning for digital communication signals. Journal of Communications and Networks, 2021, 23 (2): 192 - 200.

[12] i, H., Xu, L., Chen, X. Transformer-based radio modulation mode recognition in wireless communications. IEEE Transactions on Wireless Communications, 2022, 21 (3): 1430 - 1440.

[13] Xing, L., Xu, W., Zhang, Y. A modulation recognition algorithm of communication signals based on transformer and self-attention mechanism. IEEE Access, 2021, 9: 12342 - 12355.

[14] Cao Zhigang, Qian Yasheng. Modern Communication Principles. Beijing: Tsinghua University Press, 1992.