Online Automatic Modulation Classification Based on Distributional Signal Representation

Xinpeng Li[1,2,⋆], Zile Jiang[1,⋆], Kai Ming Ting[1,2], and Ye Zhu[3]

[1] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
lixp@lamda.nju.edu.cn, zilerion99@gmail.com, tingkm@nju.edu.cn
[2] School of Artificial Intelligence, Nanjing University, Nanjing 210023, China
[3] School of Information Technology, Deakin University, Geelong 3125, VIC, Australia
ye.zhu@ieee.org

Abstract. Automatic Modulation Classification (AMC), as a crucial technique in modern non-cooperative communication networks, plays a key role in various civil and military applications. However, existing AMC methods based on deep learning are often excessively complex and restricted to batch-mode processing due to their high computational overhead. Furthermore, they usually rely on indirect signal representations that can compromise performance. To address these issues, this paper introduces a new online AMC scheme based on a direct and lossless distributional representation of signals. It works well in online settings under realistic time-varying channel conditions. Through extensive experiments in online settings, we demonstrate the effectiveness of the proposed classifier. Our results indicate that the proposed approach outperforms existing baseline models, including two advanced deep learning classifiers. Moreover, it distinguishes itself as the first online classifier for AMC with linear time complexity, which marks a significant efficiency boost for real-time applications.

Keywords: Automatic modulation classification · Signal representation · Distributional kernel · Online learning · Time-varying channel conditions.

1 Introduction

With the rapid development of next-generation communication technologies, automatic modulation classification (AMC) has come to play a significant role in modern communication networks. On the receiver side, AMC monitors and automatically recognizes the modulation format of incoming signals in real-time, ensuring correct demodulation and reliable data recovery [28].

While deep learning (DL)-based classifiers have emerged as a popular paradigm for AMC, they are constrained by significant drawbacks that often compromise their effectiveness in practical applications. First, most DL-based models rely

⋆ These authors contributed equally to this work.

on indirect representations of signals, which require a separate data conversion step prior to classification that could potentially result in information loss. Second, their complex architectures make hyperparameter tuning and training a resource-intensive and time-consuming process [2]. Finally, their reliance on offline batch training makes them unable to adapt to realistic time-varying channel conditions, leading to significant performance degradation when a mismatch occurs between training and deployment environments. While integrating an online update mechanism could compensate for this, it introduces further complexity.

To overcome these issues, this paper introduces IDK-OGD, a novel online scheme for AMC. The main contributions are:

1. We introduce an insight that the constellation diagram of a baseband signal sequence can be losslessly represented as a distribution. This distributional representation of signals enables the use of a distributional kernel for both extracting knowledge from signals as high-dimensional embeddings and measuring similarity between signals. It offers a fundamentally new paradigm for understanding and extracting information from signals in AMC tasks.
2. We propose IDK-OGD, an online AMC scheme that integrates the Isolation Distributional Kernel (IDK) [19] to embed the distribution exhibited in a constellation diagram with the Online Gradient Descent (OGD) classifier [7,18]. IDK-OGD is the first kernel-based classifier with linear time complexity that works well in online AMC environments where there is a mismatch between the channel conditions of the training and testing data.
3. Through extensive experiments comparing IDK-OGD against advanced DL-based methods and other baselines, we demonstrate its effectiveness and practicality on both synthetic and real-world datasets, particularly under challenging mismatched and time-varying channel conditions.

2 Related work

Over the past few years, with the remarkable advancement of DL, numerous DL-based AMC methods have been proposed [8,22,4]. These DL-based methods are generally categorized into four types based on their signal representation strategy, *i.e.*, engineered feature, image, sequence, and combined representations [14]. We analyze these representations below to examine how prior research has utilized different techniques to represent signals.

Engineered feature representations such as high-order cumulants can be calculated from the original baseband signals and are prevalent in DL-based AMC frameworks [6,21]. However, these features require manual extraction tailored to the candidate modulation formats and channel conditions, limiting their flexibility. *Image representation*: Inspired by the success of DL in processing images, some researchers have proposed converting signals into fixed-resolution images to facilitate the application of image-based DL methods in AMC [13,17]. While such methods benefit from the powerful capability of DL in extracting knowledge from images, the image conversion process inevitably introduces noise and

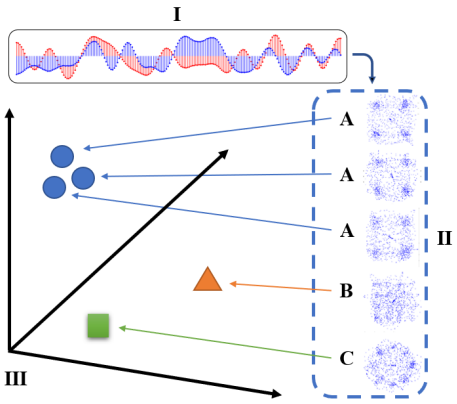


Fig. 1. Sketch of distributional signal representation for AMC (different letters, *i.e.*, A, B & C, stand for different modulation formats).

causes information loss. *Sequence representation*: these methods operate on sequences: either the raw I/Q baseband signal sequences themselves [12,3], or some transformed versions of these signals [15,5]. The efficacy of these methods hinges on the model's ability to extract robust temporal features from the sequences. *Combined representations* try to exploit the advantages of multiple signal representations [23,27]. However, whether these combinations yield a synergistic effect, or simply increase computational overhead, remains poorly understood.

To summarize, most current DL-based methods rely on indirect representations that incur information loss during the data conversion. In contrast, we propose a direct, lossless distributional signal representation in the next section.

3 Distributional Signal Representation

In this paper, we focus on digital amplitude-phase (A/P) modulated signals, *e.g.*, amplitude-shift keying (ASK), phase-shift keying (PSK), amplitude and phase-shift keying (APSK), and quadrature amplitude modulation (QAM). These modulation formats find widespread utilization in practical wireless communications and pose challenges for conventional AMC schemes.

In general, a baseband signal $r(t)$ received from a wireless channel can be expressed as:

$$r(t) = s(t) * c(t) + w(t) \tag{1}$$

where $s(t)$ denotes the transmitted baseband signal, $c(t)$ is the channel impulse response, $w(t)$ represents the additive noise, and $*$ is a convolution operator.

The transmitted A/P modulated baseband signal $s(t)$ with N symbols can be expressed as:

$$s(t) = \sum_{n=1}^{N} a_n \exp(j\varphi_n) g(t - nT_s) \tag{2}$$

and

$$a_n \exp(j\varphi_n) = a_n \cos\varphi_n + j \cdot a_n \sin\varphi_n = I_n + j \cdot Q_n \tag{3}$$

where a_n and φ_n are the amplitude and phase of the n-th symbol; $g(t)$ signifies the pulse shaping filter; T_s is the symbol duration; and I_n and Q_n denote the in-phase (I) and quadrature (Q) components of the n-th symbol, respectively.

In essence, the difference between two A/P modulation formats lies in the arrangements of A/P pairs (a_n, φ_n) of all N symbols [9].

As shown in Eq. 3, an A/P modulated signal consists of a sequence of complex data points, of which the real and imaginary components are the I and Q components of the signal, respectively. As shown in Fig. 1 from block I to II, this sequence can be seen as a scatter plot of a set of points in a two-dimensional complex space of I and Q axes, which is commonly referred to as a constellation diagram [1]. With Eq. 3, it is evident that the constellation diagram of an A/P modulated signal encompasses all the amplitude and phase information of the signal. Consequently, constellation diagrams are equivalent to the original signal sequences for AMC purposes, preserving all relevant information.

Our insight is that **a signal S in the I/Q space (or constellation diagram) can be treated as a set of independent and identically distributed (i.i.d.) points x drawn from an unknown probability distribution $\mathcal{P}_S$, *i.e.*, $x \sim \mathcal{P}_S$, without loss of information**.

With this insight, the task of AMC in identifying the unknown modulation format of a signal S can be achieved by maximizing the similarity between distribution $\mathcal{P}_S$ and the distribution $\mathcal{P}_{S_\alpha}$ of a modulation format α out of m candidate modulation formats $\{M_1, M_2, \cdots, M_m\}$. It is formally expressed as:

$$\mathrm{AMC}(S) = \underset{\alpha \in \{M_1, M_2, \cdots, M_m\}}{\arg\max} \mathcal{K}(\mathcal{P}_S, \mathcal{P}_{S_\alpha}) \tag{4}$$

where $\mathcal{K}$ is a distributional kernel which measures the similarity between two distributions; and S_α is the i.i.d. sample set of signals with modulation format α. Eq. 4 returns the modulation format α of $\mathcal{P}_{S_\alpha}$ which is most similar to $\mathcal{P}_S$.

The task is made simple by using a feature mapping derived from $\mathcal{K}$ with an injective mapping guarantee [19,10] such that signals of the same modulation format are mapped into the same region in the high-dimensional feature space, while those of different modulation formats are mapped into different regions, as shown in Fig. 1 (from II to III). This ensures a higher similarity between signal distributions sharing the same modulation format.

Note that the choice of signal representation plays a pivotal role in determining the performance of an AMC classifier. The sequence and image representations commonly utilized in DL necessitate the transformation of the original signals into specific sequential or grid-based formats as input, which may result in unnecessary information loss and the introduction of noise. In contrast, our proposed distributional signal representation is a direct approach that preserves the original signals within their native I/Q domains in a two-dimensional I/Q signal space, eliminating the need for conversion, and thus the associated information loss and potential corruption.

4 Proposed Method

Kernel methods are widely employed in classification and pattern recognition tasks. With the above insight of distributional signal representation, we propose to use a recent Isolation Distributional Kernel (IDK) [19] to measure the similarity between two signals in their distributional representation. IDK is a well-established method that has already been applied in various applications, and detailed specifications can be found in [19]. The feature mapping of IDK, which maps a distribution $\mathcal{P}_S$ (from which the set of points $S \subset \mathbb{R}^d$ is drawn) to a single point in a high-dimensional feature space known as the Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ [16], is derived via kernel mean embedding [10] using the feature mapping $\Phi(x)$ of the point-wise Isolation Kernel (IK) [20,25]:

$$\widehat{\Phi}(\mathcal{P}_S) = \frac{1}{|S|} \sum_{x \in S} \Phi(x) \tag{5}$$

The similarity between two distributions $\mathcal{P}_S$, $\mathcal{P}_T$ measured by IDK is given as:

$$\mathcal{K}(\mathcal{P}_S, \mathcal{P}_T) = \left\langle \widehat{\Phi}(\mathcal{P}_S), \widehat{\Phi}(\mathcal{P}_T) \right\rangle \tag{6}$$

A key property of IDK is its finite-dimensional, data-dependent feature mapping derived directly from a given dataset to yield better task-specific performance [20]. The derivation is highly efficient as it bypasses an explicit learning process. Furthermore, IDK offers an injective mapping guarantee, ensuring that two distributions $\mathcal{P}_S$ and $\mathcal{P}_T$ are mapped to the same point in $\mathcal{H}$ only when they are exactly identical, *i.e.*, $\| \widehat{\Phi}(\mathcal{P}_S) - \widehat{\Phi}(\mathcal{P}_T) \|_{\mathcal{H}} = 0$ if and only if $\mathcal{P}_S = \mathcal{P}_T$ [24]. This is a significant theoretical advantage, as none of the existing representations used in DL-based methods can provide a similar guarantee.

Our proposed IDK-OGD classifier is designed for online AMC by integrating IDK for the embedding of signals with a large-scale online classifier Online Gradient Descent (OGD) [7]. It builds upon IK-OGD which has proven effective in conventional data stream contexts [18]. We adapt IK-OGD for the AMC task by substituting the point-wise IK with the distributional IDK to accommodate the distributional representation of I/Q baseband signals, and extending the binary classifier to a multi-class variant capable of discriminating between multiple modulation formats. This novel application is underpinned by the distributional signal representation we introduced, which makes these signals amenable to powerful distributional kernel-based online learning techniques.

The operational workflow of the proposed IDK-OGD classifier is as follows. Suppose there are m candidate modulation formats denoted by $M_1, M_2, \cdots, M_m$. IDK-OGD is first initialized using a batch of training data to construct m distinct IDK feature mappings. Each feature mapping, $\widehat{\Phi}_j$, is derived exclusively from the training signals corresponding to a single modulation format M_j. During the classification stage, an incoming I/Q baseband signal is passed through all m feature mappings, and a corresponding set of m OGD classifiers generates a decision score for each class. The final prediction is determined by selecting the

Algorithm 1 IDK-OGD

Input:

m - The number of modulation formats
$\widehat{\Phi}_j$ - IDK feature mapping of modulation format j
η - Learning rate

Output:

Predicted modulation format $\widehat{c}_i$ for each signal S_i

1: Initialize weight vector $\mathbf{w}_j$ to $\mathbf{0}$ for $j = 1, \ldots, m$
2: **while** a new batch of signals $\mathcal{B}$ arrives **do**
3: **for** $i = 1 : |\mathcal{B}|$ **do**
4: **for** $j = 1 : m$ **do**
5: $g_{i,j} = \left\langle \mathbf{w}_j, \widehat{\Phi}_j(\mathcal{P}_{S_i}) \right\rangle$
6: **end for**
7: $\widehat{c}_i = \arg\max_j \ g_{i,j}$
8: **end for** ▷ Classification stage ends
9: **if** ground-truth labels c_i for $S_i \in \mathcal{B}$ are available **then**
10: **for** $i = 1 : |\mathcal{B}|$ **do**
11: $k_i = \begin{cases} 1, & \text{if } c_i = \widehat{c}_i \\ -1, & \text{if } c_i \neq \widehat{c}_i \end{cases}$
12: **for** $j = 1 : m$ **do**
13: $\mathbf{w}_j = \mathbf{w}_j - \eta \nabla L(g_{i,j}; k_i) k_i \widehat{\Phi}_j(\mathcal{P}_{S_i})$
▷ ∇L is the gradient of the hinge loss function L
14: **end for**
15: **end for**
16: **end if** ▷ Online model update stage ends
17: **end while**

modulation format with the highest score. Subsequently, all OGD classifiers are updated simultaneously if the ground-truth labels for the signals are available. If labels are unavailable, the model suspends the update mechanism and functions as a robust static classifier.

Algorithm 1 provides the pseudo-code for the proposed IDK-OGD classifier, detailing the classification and online update stages. It is worth noting that the calculations for each modulation format (as shown in the for-loops at lines 4-6 and 12-14) are independent, allowing IDK-OGD to be implemented using a parallel architecture, which significantly improves its computational efficiency.

5 Experiments

5.1 Experimental Settings

Our experiments utilize the open-source RadioML2018.01A dataset [11] and synthetic datasets generated via MATLAB Communications Toolbox. The RadioML2018.01A dataset offers diverse 1024-point I/Q signals with various channel effects such as carrier frequency offset, multipath fading, sample rate offset,

etc., which make the task more challenging. However, it provides labels exclusively for signal-to-noise ratio (SNR), which quantifies additive white Gaussian noise (AWGN), lacking labels for other impairments such as phase noise or I/Q amplitude imbalance. Our synthetic data was generated to enable controlled studies under these effects. The evaluation focuses on ten modulation formats (4ASK, 8ASK, BPSK, QPSK, 8PSK, 16APSK, 32APSK, 16QAM, 32QAM, and 64QAM) for their widespread use and the challenges they pose to AMC.

Note that most existing AMC methods operate under the assumption that training and testing data share identical and fixed channel conditions. However, this assumption is often violated in real-world applications where channel conditions are time-varying and key parameters like SNR cannot be accurately estimated in real-time. Therefore, our experiments are designed to assess classifier performance under mismatched conditions, where the channel conditions of the testing data intentionally deviate from those of the training data. To simulate real-time data arrival, the testing set is organized into sequential batches of 100 signal samples, which are processed by the classifiers one by one; and to simulate time-varying channel conditions, every 10 batches are grouped together to share the identical channel conditions, while different groups of signal batches may exhibit different channel conditions.

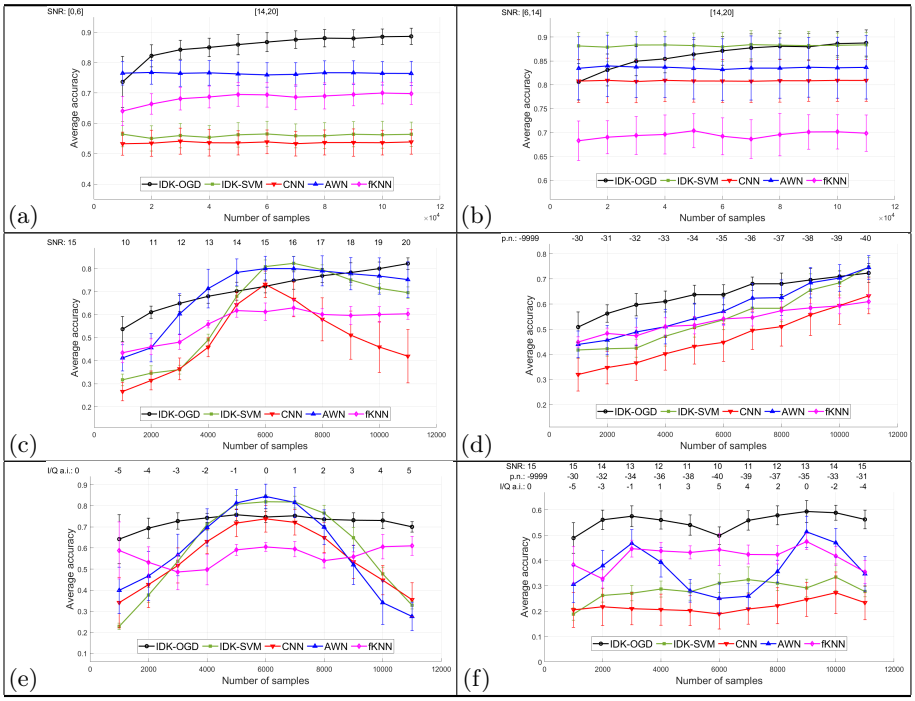
We compare the proposed IDK-OGD classifier against various baselines for AMC, including fKNN [29], an online feature-based classifier using engineered high-order moments, and two advanced DL-based models: a Convolutional Neural Networks (CNN) using image for signal representation [17] [4] and a Adaptive Wavelet Network (AWN) with sequence representations [26]. We also compare with IDK-SVM, a variant that pairs the IDK representation with a conventional, non-online Support Vector Machine (SVM) classifier, allowing for a direct evaluation of our method's effectiveness due to an online classifier. These methods illustrate three distinct adaptation strategies: incremental online update (IDK-OGD), retraining on cumulative data (fKNN), and static batch-learning (CNN, AWN, IDK-SVM) [5] .

The first two experiments are conducted on the RadioML2018.01A dataset. In both experiments, the testing sets are maintained at high SNR, while the training sets exhibit low SNR in the first case and medium SNR in the second. In the latter configuration, the training SNR range partially overlaps with that of the testing set. Furthermore, we design four additional experiments with mismatched training and testing channel conditions and different channel impairments, including AWGN, phase noise, I/Q amplitude imbalance, and a combination of all the three impairments. These experiments are conducted on synthetic datasets

[4] While the CNN baseline [17] claims to have an online version involving a retraining process, only its no-retraining version is accessible. We therefore adopted the static version for comparison.

[5] For the IDK component, the hyperparameters are set to $\psi = 128$ and $t = 75$. Both OGD and SVM utilize a linear kernel, with the box constraint for SVM set to $C = 980$. For the fKNN baseline, the number of neighbors is set to $k = 15$. The hyperparameters for the CNN and AWN models are tuned according to the specifications provided in their respective original works [17,26].

Table 1. Experimental results across six scenarios on the RadioML2018.01A and synthetic datasets, including: (a) mismatched and (b) overlapping SNR on RadioML2018.01A; and mismatched (c) SNR, (d) phase noise, (e) I/Q amplitude imbalance, and (f) all three impairments combined on synthetic datasets. In each plot, the parameters for the training and testing channel conditions are specified at the top.



to facilitate controlled studies of the performance of the classifiers under different mismatched time-varying channel conditions. For initialization, all classifiers are trained on the identical dataset, of which the size matches the total testing set (110,000 samples in the first two experiments and 11,000 samples in the others).

5.2 Experimental Results

The experimental results are presented in Table 1, where classifier performance is evaluated based on the average prediction accuracy over 10 trials. Each subfigure illustrates a distinct experimental scenario, with the parameters listed at the top specifying the fixed training condition (leftmost) and the time-varying testing channel conditions. The curves below plot the corresponding accuracy of each method as the testing environment changes. Specifically, subfigures (a)-(b) show results on the RadioML2018.01A dataset for mismatched and overlapping SNR, while subfigures (c)-(f) use synthetic data to assess performance under mismatched SNR, phase noise, I/Q imbalance, and a combination of them all.

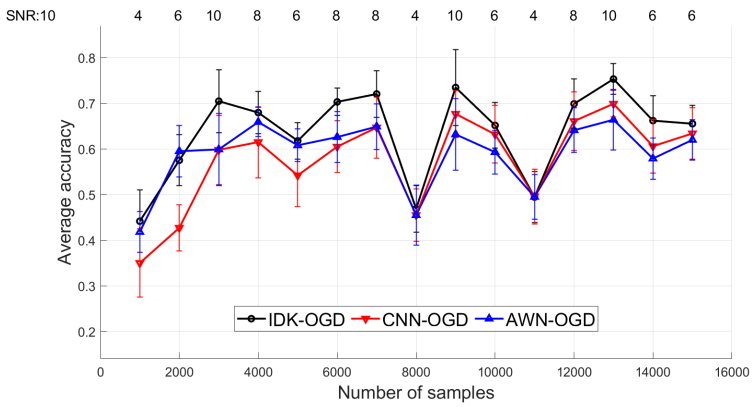


Fig. 2. The performance of IDK-OGD, CNN-OGD and AWN-OGD.

The primary finding from our experiments is the superior performance of the proposed IDK-OGD. As shown in Table 1 (a) and (b), IDK-OGD maintains high accuracy on the RadioML2018.01A dataset under both complete and partial SNR mismatch. While its initial accuracy is slightly below that of AWN in both cases, its online update mechanism enables rapid adaptation to new testing channel conditions. By learning from incoming signal batches, IDK-OGD quickly surpasses all other methods. This robustness is further validated on the synthetic datasets (subfigures (c)-(f)). In these scenarios, the performance of all baseline methods degrades significantly in the presence of mismatched impairments, whereas IDK-OGD consistently outperforms them by a large margin.

The batch-learning baselines (IDK-SVM, CNN, and AWN) exhibit similar performance trends. They achieve higher accuracy than IDK-OGD only when the testing channel conditions remain consistent with the initial training data. However, their performance degrades substantially as the mismatch between the environments increases, due to their lack of an adaptation mechanism. While the retraining-based method fKNN shows certain robustness to channel mismatch, its performance remains inconsistent and is significantly inferior to IDK-OGD across all experiments.

5.3 Ablation Study

Here we empirically demonstrate the superiority of our distributional representation over existing signal representations via an ablation study where we compare the performance of IDK-OGD with CNN-OGD and AWN-OGD (online-enabled DL-based models produced by combining the embeddings learned by the DL-based CNN/AWN with the OGD classifier). CNN-OGD utilizes the output of the pooling layer (*i.e.*, the result of the representation learning) in CNN, while AWN-OGD employs the latent embeddings prior to the final fully connected layers (*i.e.*, before classification) in AWN, for OGD classification. The result given in Fig. 2 shows that IDK-OGD almost always outperforms both CNN-OGD and

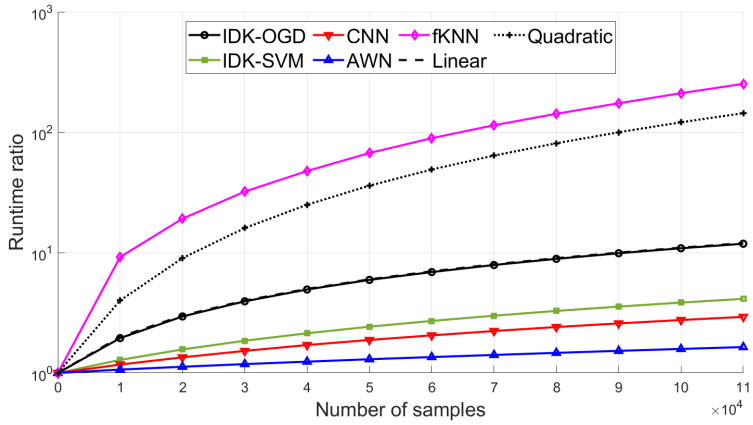


Fig. 3. Runtime ratio for the classifiers.

AWN-OGD. This highlights the power of the distributional representation via IDK and its advantage over other signal representations used in the two DL-based methods (*i.e.*, the sequence and image representations).

5.4 Runtime Comparison

Fig. 3 shows the ratio of the total runtime to the training time, where the IDK-OGD curve overlaps with the linear curve, demonstrating the superior scalability of IDK-OGD. Its runtime grows linearly with the input data size n, aligning with its complexity of $O(nmLt\psi)$. The other terms, *i.e.*, the number of modulation formats m, signal length L, and the IDK parameters t and ψ, are constants. In contrast, fKNN's retraining process results in quadratic complexity. While the batch-learning baselines (IDK-SVM, CNN, AWN) have rapid inference, they are non-adaptive, and retraining them would be computationally infeasible. The linear-time online update of IDK-OGD is, to our knowledge, unique among AMC methods, establishing its efficiency for real-time deployment.

6 Conclusions

In this paper, we introduce a novel distributional signal representation that bridges the conceptual gap between powerful distributional measures such as distributional kernels and AMC tasks.

Building on this, we propose the IDK-OGD classifier and conduct extensive evaluation to demonstrate that it is highly effective and robust as an online AMC method under realistic time-varying channel conditions, especially when a mismatch of environment occurs. Two key factors contribute to its success: a powerful distributional kernel for the embedding of signals, and a robust online learning algorithm that enables rapid adaptation. IDK-OGD is also the first online classifier with linear time complexity for AMC, as far as we know.

Looking forward, we believe the distributional signal representation presented in this work has far-reaching implications. It paves the way for a new class of solutions for AMC that has remained largely unexplored, encouraging the broader application of distributional measures, both existing and new, to this important domain.

Acknowledgments. This project is supported by the State Key Laboratory for Novel Software Technology at Nanjing University (Grant No. KFKT2024A01).

References

1. Carvalho, N.B., Schreurs, D.: Microwave and Wireless Measurement Techniques. The Cambridge RF and Microwave Engineering Series, Cambridge University Press, Cambridge, England (2013)
2. Dai, Y., Gao, X., Jing, K., Tian, B.: High-efficiency modulation classification with temporal-frequency analysis based on multi-channel filter bank. In: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2025)
3. Dong, B., Liu, Y., Gui, G., Fu, X., Dong, H., Adebisi, B., Gacanin, H., Sari, H.: A lightweight decentralized-learning-based automatic modulation classification method for resource-constrained edge devices. IEEE Internet of Things Journal **9**(24), 24708–24720 (2022)
4. Hu, X., Gao, G., Li, B., Wang, W., Ghannouchi, F.M.: A novel lightweight grouped gated recurrent unit for automatic modulation classification. IEEE Wireless Communications Letters **13**(8), 2135–2139 (2024)
5. Khan, F.N., Zhong, K., Al-Arashi, W.H., Yu, C., Lu, C., Lau, A.P.T.: Modulation format identification in coherent receivers using deep machine learning. IEEE Photonics Technology Letters **28**(17), 1886–1889 (2016)
6. Lee, S.H., Kim, K.Y., Kim, J.H., Shin, Y.: Effective feature-based automatic modulation classification method using DNN algorithm. In: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (2019)
7. Lu, J., Hoi, S.C., Wang, J., Zhao, P., Liu, Z.Y.: Large scale online kernel learning. Journal of Machine Learning Research **17**(47), 1–43 (2016)
8. Mao, Q., Hu, F., Hao, Q.: Deep learning for intelligent wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials (2018)
9. Mao, Y., Dong, Y.Y., Sun, T., Rao, X., Dong, C.X.: Attentive siamese networks for automatic modulation classification based on multitiming constellation diagrams. IEEE Transactions on Neural Networks and Learning Systems (2021)
10. Muandet, K., Fukumizu, K., Sriperumbudur, B., Schölkopf, B.: Kernel mean embedding of distributions: A review and beyond. Foundations and Trends in Machine Learning **10 (1–2)**, 1–141 (2017)
11. O'Shea, T., West, N.: Radio machine learning dataset generation with GNU radio. Proceedings of the GNU Radio Conference **1**(1) (2016)
12. O'Shea, T.J., Corgan, J., Clancy, T.C.: Convolutional radio modulation recognition networks. In: Engineering Applications of Neural Networks: 17th International Conference, EANN 2016, Aberdeen, UK, September 2-5, 2016, Proceedings 17. pp. 213–226. Springer (2016)

13. Peng, S., Jiang, H., Wang, H., Alwageed, H., Zhou, Y., Sebdani, M.M., Yao, Y.D.: Modulation classification based on signal constellation diagrams and deep learning. IEEE Transactions on Neural Networks and Learning Systems (2019)
14. Peng, S., Sun, S., Yao, Y.D.: A survey of modulation classification using deep learning: Signal representation and data preprocessing. IEEE Transactions on Neural Networks and Learning Systems **33**(12), 7020–7038 (2022)
15. Rajendran, S., Meert, W., Giustiniano, D., Lenders, V., Pollin, S.: Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. IEEE Transactions on Cognitive Communications and Networking (2018)
16. Schölkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press (2001)
17. Teng, C.F., Chou, C.Y., Chen, C.H., Wu, A.Y.: Accumulated polar feature-based deep learning for efficient and lightweight automatic modulation classification with channel compensation mechanism. IEEE Transactions on Vehicular Technology **69**(12), 15472–15485 (2020)
18. Ting, K.M., Wells, J.R., Washio, T.: Isolation kernel: The X factor in efficient and effective large scale online kernel learning. Data Mining and Knowledge Discovery **35**, 2282–2312 (2021)
19. Ting, K.M., Xu, B.C., Washio, T., Zhou, Z.H.: Isolation distributional kernel: A new tool for point & group anomaly detections. IEEE Transactions on Knowledge and Data Engineering (2021)
20. Ting, K.M., Zhu, Y., Zhou, Z.H.: Isolation kernel and its effect on SVM. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2329–2337 (2018)
21. Wang, H., Ding, W., Zhang, D., Zhang, B.: Deep convolutional neural network with wavelet decomposition for automatic modulation classification. In: 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA) (2020)
22. Wang, Y., Gui, G., Ohtsuki, T., Adachi, F.: Multi-task learning for generalized automatic modulation classification under non-Gaussian noise with varying SNR conditions. IEEE Transactions on Wireless Communications **20**(6), 3587–3596 (2021)
23. Wang, Y., Liu, M., Yang, J., Gui, G.: Data-driven deep learning for automatic modulation recognition in cognitive radios. IEEE Transactions on Vehicular Technology **68**(4), 4074–4077 (2019)
24. Wang, Z.J., Zhu, Y., Ting, K.M.: Distribution-based trajectory clustering. In: 2023 IEEE International Conference on Data Mining (ICDM) (2023)
25. Xu, B.C., Ting, K.M., Zhou, Z.H.: Isolation set-kernel and its application to multi-instance learning. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 941–949 (2019)
26. Zhang, J., Wang, T., Feng, Z., Yang, S.: Toward the automatic modulation classification with adaptive wavelet network. IEEE Transactions on Cognitive Communications and Networking **9**(3), 549–563 (2023)
27. Zhang, Z., Wang, C., Gan, C., Sun, S., Wang, M.: Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD. IEEE Transactions on Signal and Information Processing over Networks (2019)
28. Zheng, Q., Tian, X., Yu, L., Elhanashi, A., Saponara, S.: Recent advances in automatic modulation classification technology: Methods, results, and prospects. International Journal of Intelligent Systems **2025**(1), 4067323 (2025)
29. Zhu, Z., Nandi, A.K.: Automatic modulation classification: principles, algorithms and applications. John Wiley & Sons, Hoboken, New Jersey (2015)